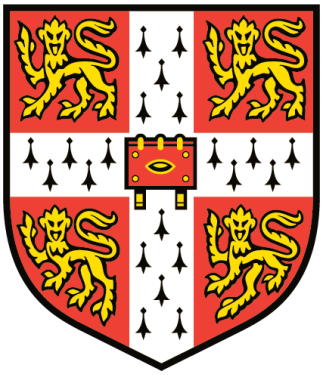
Development of *in vitro* iCLIP techniques to study spliceosome remodelling by RNA helicases



LISA MARIA STRITTMATTER

MRC Laboratory of Molecular Biology

Newnham College

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

July 2018

Title: Development of *in vitro* iCLIP techniques to study spliceosome remodelling by RNA helicases
Name: Lisa Maria Strittmatter

Thesis summary

Pre-mRNA (precursor messenger RNA) splicing is a fundamental process in eukaryotic gene expression. In order to catalyse the excision of the intervening intronic sequence between two exons, the spliceosome is assembled stepwise on the pre-mRNA substrate. This ribonucleoprotein machine is extremely dynamic: both its activation and the progression through the catalytic stages require extensive compositional and structural remodelling.

The first part of this thesis aims at understanding how the spliceosome is activated after assembly. When this work was started, the GTPase Snu114 was thought to activate the helicase Brr2 to unwind the U4/U6 snRNA duplex, which ultimately leads to the formation of the spliceosome active site. To explore the role of Snu114, a complex built from Snu114 and a part of Prp8 was expressed and analysed in its natural context, bound to U5 snRNA. However, before I was able to obtain highly diffracting crystals, the structure of Snu114 was determined in the context of a larger spliceosomal complex by electron cryo-microscopy by competitors. Regardless, the role of Snu114 in spliceosome activation remains elusive. In a short section of this thesis, genetic and biochemical analysis suggest Snu114 to be a pseudo-GTPase, precluding a role for Snu114-catalyzed GTP hydrolysis in activation.

The second and larger part of the thesis describes the development of a novel, biochemical method to analyse spliceosome remodelling events that are caused by the eight spliceosomal helicases. Purified spliceosomes assembled on a defined RNA substrate are analysed by UV crosslinking and next-generation sequencing, which allows for the determination of the RNA helicase binding profile at nucleotide resolution. *In vitro* spliceosome iCLIP (individual-nucleotide resolution UV crosslinking and immunoprecipitation) was initially developed targeting the helicase Prp16 bound to spliceosomal complex C. The obtained binding profile shows that Prp16 contacts the intron, about 15 nucleotides downstream of the branch in the intron-lariat intermediate. Our finding supports the model of Prp16 acting at a distance to remodel the RNA and protein interactions in the catalytic core and thereby it promotes the transition towards a conformation of the spliceosome competent for second step catalysis. Control experiments, which locate SmB protein binding to known Sm sites in the spliceosomal snRNAs, validated the method. Preliminary results show that *in vitro* spliceosome iCLIP can be adapted to analyse additional spliceosomal helicases such as Prp22. Finally, I performed initial experiments that give promising directions towards time-resolved translocation profiles of helicases Brr2 and Prp16.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

This dissertation does not exceed the prescribed word limit of 60,000 words as defined by the Degree Committee Biology.

Table of contents

List of Figures

List of Tables

Abbreviations, variables and units

3'SS	3' splice site
5'SS	5' splice site
5-FOA	5-fluoro orotic acid
Å	angstrom
aa	amino acid
ADP	adenosine diphosphate
AP	alkaline phosphatase
ATP	adenosine triphosphate
BPS	branch point sequence
brA	branchpoint adenosine
C	Celcius
CBP	calmodulin-binding protein
CLIP	cross-linking and immunoprecipitation
cryoEM	electron cryo-microscopy
Da	dalton
DMSO	dimethyl sulfoxide
DN	dominant negative
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E. coli	*Escherichia coli*
eCLIP	enhanced CLIP
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid
EMSA	electrophoretic mobility shift assays
g	gram
g	standard gravity
GAP	GTPase activating proteins
GEF	GTP exchange factor
GTP	guanosine triphosphate
h	hour
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HITS-CLIP	high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation
HPLC	high-performance liquid chromatography
iCLIP	individual nucleotide resolution UV cross-linking and immunoprecipitation
IGV	integrative genomics viewer
ILS	intron-lariat spliceosome
IR	infrared
J	joule
L	litre
m	meter
M	molar
MALS	multi-angle light scattering
MBP	maltose-binding protein
min	minute
mol	mole
NGS	next-generation sequencing
Ni-NTA	nickel nitrilotriacetic acid
nt	nucleotides
NTC	nineteen complex
NTP	nucleoside triphosphate
NTR	nineteen related complex
PAGE	polyacrylamide gel electrophoresis
PAR-CLIP	photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation
PBS	phosphate-buffered saline
PCI	phenol:chloroform:isoamyl alcohol
PCR	polymerase chain reaction
PEG	polyethylene glycol
PNK	polynucleotide kinase
pre-mRNA	precursor messenger RNA
RBP	RNA binding protein

RNA	ribonucleic acid
RNP	ribonucleoprotein
RT	reverse transcription
RU	relative units
S. cerevisiae	*Saccharomyces cerevisiae*
SDS	sodium dodecyl sulphate
sec	second
SL	stem loop
snRNA	small nuclear RNA
snRNP	small nuclear RNP
SS	splice site
TBE	Tris-borate-EDTA
TEV	tobacco etch virus
Tris	tris(hydroxymethyl)amino-methane
U	units
UMI	unique molecular identifier
UTR	untranslated region
UV	ultraviolet
UV	ultraviolet
WT	wild-type
XNP	xanthine nucleotide
XTP	xanthine triphosphate
YEPD	yeast extract peptone dextrose
β-ME	β-mercaptoethanol

mutant proteins are denoted by the one-letter amino acid (aa) code for point mutations, e.g. Prp16-G378A; dominant negative proteins are denoted by protein-DN, e.g. Prp16-DN; WT proteins are sometimes denoted by protein-WT to distinguish from mutant proteins

recombinant proteins are denoted by an additional r in chapters 3, 4, and 5, e.g. rPrp16

protein tags are denoted by tag-protein for N-terminal tags and as protein-tag for C-terminal tags, e.g. CBP-Prp22 and Prp16-3xFLAG

protein oligomers are denoted by : between the components, e.g. Snu114:Prp8

protein domains are denoted by protein-domain, e.g. Prp8-N for the N-terminal domain of Prp8

protein fragments are denoted by protein_x-y where x and y give the position of the first and last aa, respectively, e.g. Prp8_1-770

1 Introduction

1.1 Introducing the spliceosome, a dynamic RNP enzyme

Pre-mRNA splicing is a crucial step in eukaryotic gene expression. Most eukaryotic genes are transcribed as precursors that contain intronic sequences (introns) interrupting the coding regions (exons) in a mosaic structure. In higher eukaryotes, pre-mRNA splicing enables cell-type specific expression of many proteins and enhances proteomic diversity via alternative splicing (Roberts and Smith 2002; Hallegger, Llorian, and Smith 2010). In addition to expanding the proteome from single genes, splicing has been reported to be crucial for: efficient transcription (Furger et al. 2002), mRNA export (Valencia, Dias, and Reed 2008), RNA localisation (Le Hir et al. 2001), transcript stability and efficient translation (Moore and Proudfoot 2009). Furthermore, the excised introns are involved in the biogenesis of noncoding RNAs (Pawlicki and Steitz 2010).

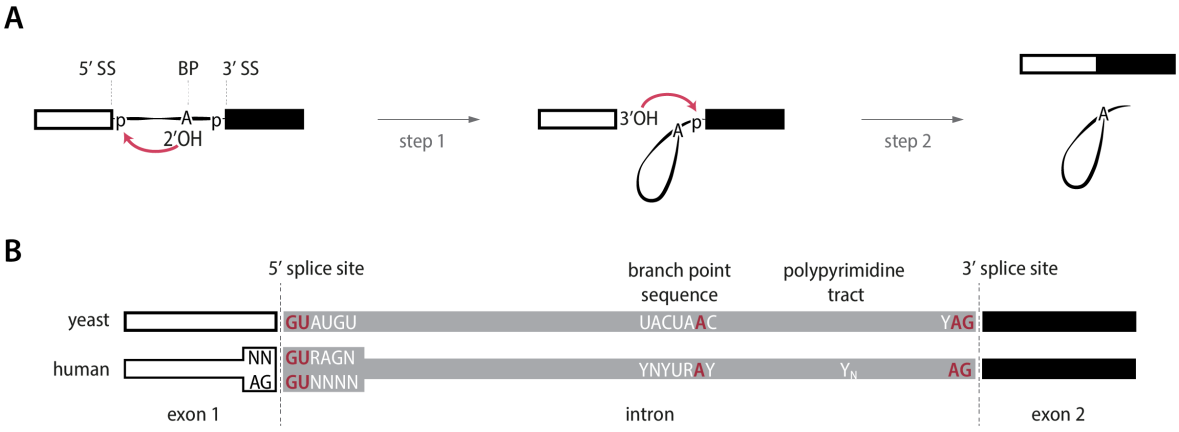


Figure 1-1: Two step splicing mechanism and consensus sequences. (**A**) Introns are excised in two consecutive transesterification reactions. Step 1 or branching generates the intron lariat-intermediate and a free 5′ exon. Step 2 or exon ligation results in the joined exons and the excised intron lariat. The nucleophilic attacks are indicated with red arrows. (**B**) Consensus sequences of the three reactive groups in *Saccharomyces cerevisiae* (*S. cerevisiae*) (top) and human (bottom). The 5′SS is located at the 5′-end of the intron and the 3′SS at the 3′-end of the intron, while the branch point sequence (BPS) is found 18-40 nucleotides upstream of the 3′SS (Reed 1989; Spingola et al. 1999; Sheth et al. 2006). Strictly conserved residues are coloured in red, including the brA within the BPS. N represents any nucleotide, R any purine, and Y any pyrimidine. Most human introns contain a stretch of 15 – 20 pyrimidines (polypyrimidine tract).

An intron that separates two exons is excised from the pre-mRNA by two consecutive SN_2-type trans-esterification reactions (**Figure 1-1 A**) (Sharp 1987). In the first reaction, named branching,

the oxygen that forms the 2'-hydroxyl group of a conserved adenosine (branchpoint adenosine, or brA) serves as a nucleophile to attack the phosphate at the 5' splice site (5'SS). A branched intron-lariat intermediate is formed, while the 5'-exon leaves with a free 3'-hydroxyl group. In the second reaction, named exon ligation, this newly generated hydroxyl group conducts a subsequent nucleophilic attack to form a bond with the phosphate at the 3' splice site (3'SS). The spliced RNA is formed, and the branched intron-lariat remains as the leaving group. All three reacting nucleotides are located in conserved regions of the intron. The human sequences are less conserved compared to yeast (**Figure 1-1 B**).

Although the splicing mechanism is rather simple from the chemical point of view, a plethora (~150) of proteins and several RNA components are required to interact in a highly ordered manner with the substrate pre-mRNA to perform the reactions in living cells (Zhou, Licklider, et al. 2002). The formed multi-megadalton ribonucleoprotein (RNP) machinery, called the spliceosome, is built from highly conserved components across eukaryotes (Wahl, Will, and Luhrmann 2009).
Most of our current knowledge about the spliceosome's composition and the splicing mechanism is derived from studies in budding yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) or human HeLa cell. Since the work in this thesis was exclusively performed with yeast components, the following introduction focuses on the yeast spliceosome.

Pre-mRNA splicing happens co-transcriptionally in the nucleus (Wallace and Beggs 2017). The spliceosome forms *de novo* on each pre-mRNA substrate by sequential joining of the so-called U snRNPs (small nuclear RNP), which are the main building blocks of the spliceosome (Lerner and Steitz 1979; Bindereif and Green 1987; Tardiff and Rosbash 2006). Each U snRNP consist of a unique small nuclear RNA (snRNA) bound by Sm or LSm proteins and several specific proteins. Transacting factors are recruited to remodel the spliceosome, which leads to activation (1) and to the catalytic stages during which the two splicing reactions occur (2). After release of the ligated exons, the spliceosome is disassembled (3) and remaining building blocks can be recycled for a new round of splicing (Wahl, Will, and Luhrmann 2009) (**Figure 1-2**).

(1) First, the U1 snRNP assembles onto the pre-mRNA substrate as the 5' region of U1 snRNA forms specific contacts by base pairing to the 5'SS (Bindereif and Green 1987; Siliciano and Guthrie 1988), while the branch point binding protein (BBP) recognises the branch point sequence (BPS). Prp5-mediated remodelling subsequently replaces the BBP with U2 snRNP whose RNA component forms base pairs with the BPS (Perriman and Ares 2010). This step requires additional remodelling by Sub2 (Fleckner et al. 1997). Overall, the association of

U1 and U2 snRNP with the pre-mRNA substrate leads to the formation of the pre-spliceosomal complex A. Next, U4/U6.U5 tri-snRNP, in which U4 and U6 snRNA are extensively base-paired, is recruited as a pre-assembled complex to form the complex pre-B (Bai et al. 2018). 5'SS transfer from U1 snRNA to U6 snRNA is mediated by Prp28 (Staley and Guthrie 1999), generating the pre-catalytic complex B (Plaschka, Lin, and Nagai 2017). Spliceosome activation happens through the release of U4 snRNP and U6 proteins, which allows the spliceosomal RNA components to fold into the catalytic core (Madhani and Guthrie 1992; Newman and Norman 1992; Sun and Manley 1995). This rearrangement comes along with the association of the nineteen and nineteen related complexes (NTC, NTR) (Chan et al. 2003) and is initiated by Brr2, which unwinds the U4/U6 snRNA duplex (Raghunathan and Guthrie 1998; Laggerbauer, Achsel, and Luhrmann 1998). For the purpose of this thesis, spliceosome activation will be defined as the transition between complex B and complex B^{act} (Yan et al. 2016a). The time point when U1 snRNP is completely released from the spliceosome is unclear and might happen during activation or earlier (compare dashed outline in **Figure 1-2**). The literature has suggested a role for Snu114 in the activation step, which will be discussed in detail in the introductory section **1.2**.

(2) To allow the first step of splicing, Prp2 initiates a remodelling event that brings the 5'SS and the brA into close proximity (Warkocki et al. 2009). Binding sites for the branching factors Cwc25 and Yju2 are generated, while SF3a/b complexes that are part of U2 snRNP are released (Chiu et al. 2009; Schneider et al. 2015). After the reaction, spliceosomal complex C is formed, harbouring the cleaved 5'-exon and the intron-lariat-3'-exon intermediate (Wan et al. 2016; Galej et al. 2016). Prp16 joins the complex, promotes release of the branching factors and binding of the exon ligation factors Slu7 and Prp18, which in turn leads to docking of the 3'SS into the active site of the spliceosome (complex C* (Yan et al. 2016b; Fica et al. 2017)).

(3) After the second step of splicing, yielding complex P (Bai et al. 2017; Liu et al. 2017; Wilkinson et al. 2017), Prp22 mediates the release of the ligated exons. Subsequently, the intron-lariat-spliceosome (ILS) (Wan et al. 2017) is disassembled into free U2, U5, and U6 snRNPs by Prp43 and Brr2 and the excised intron-lariat is released (Makarov et al. 2002). This allows the snRNPs to assemble anew on another pre-mRNA substrate.

During the past three years, an increased knowledge about the spliceosome's architecture and structure was achieved by electron cryo-microscopy (cryoEM). Both catalytic reactions were confirmed to be catalysed by the same active site formed of RNA (Fica et al. 2013; Fica et al. 2014). We can learn a lot from the snapshots of the spliceosome taken during assembly, activation, catalysis

and disassembly, however, due to the dynamic nature of the spliceosomes, most processes which happen between the stable intermediates remain poorly characterised. Only recently, the dynamics of the spliceosome became a focus for single molecule studies (DeHaven, Norden, and Hoskins 2016; Hoskins et al. 2016; Semlow et al. 2016; Larson and Hoskins 2017; van der Feltz and Hoskins 2017; van der Feltz, DeHaven, and Hoskins 2018).

Most of the remodelling steps between spliceosome intermediates require energy, which is utilised through hydrolysis of nucleoside triphosphate (NTP). The next paragraphs will focus on the remodelling steps and the enzymes that initiate the rearrangements.

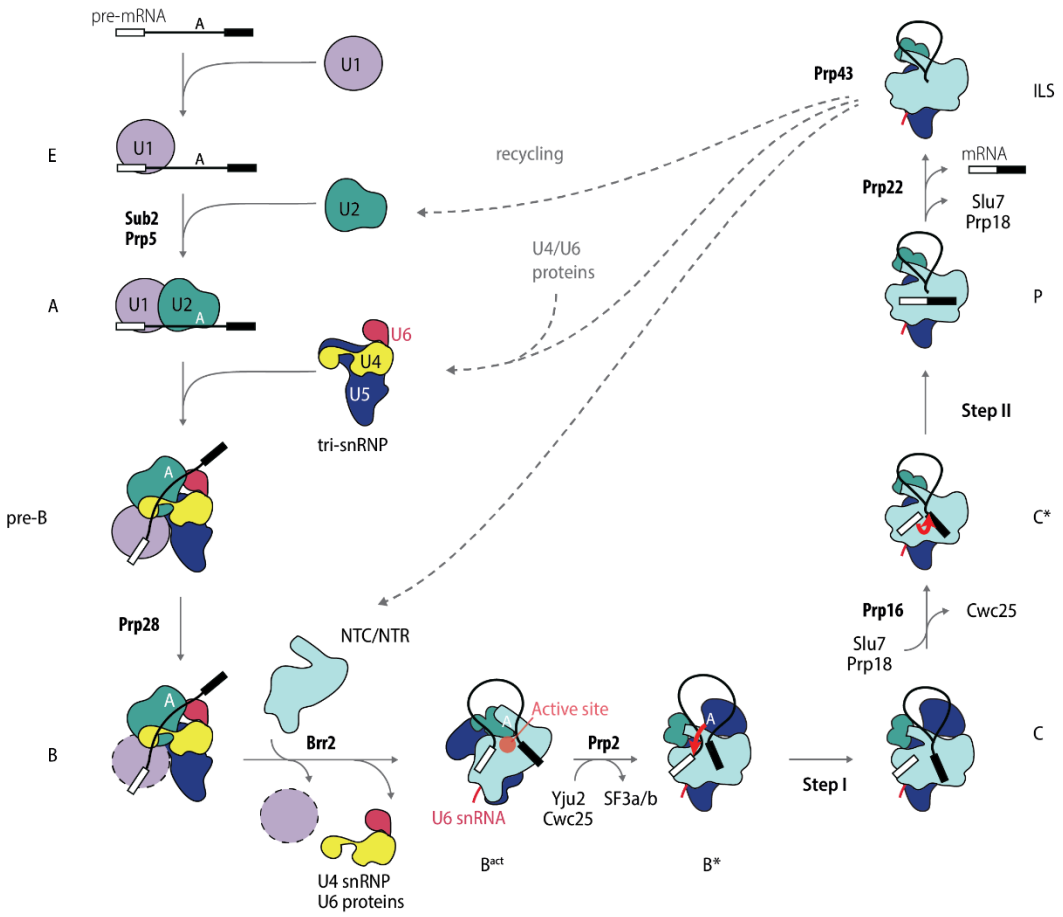


Figure 1-2: The general mechanism of pre-mRNA splicing by the spliceosome. Schematic presentation of the spliceosome intermediates which are formed during the reaction; adapted from (Wilkinson et al. 2018). Contours are based on available cryoEM structures. The pre-mRNA is depicted as: exon 1 – white box; intron – black line; exon 2 – black box. See text for description of the different stages and abbreviations. Snu114 is not indicated in the figure but discussed in the text.

1.2 Snu114 – the only spliceosomal GTPase

Snu114 stands out among all the proteins required for splicing as the only GTPase. Snu114 is an integral part of U5 snRNP and interacts with the U5 snRNA internal loop 1 (Dix et al. 1998). The

protein joins the spliceosome as part of tri-snRNP and stays associated with U5 snRNP throughout the splicing cycle.

Snu114 is homologous (26% identity, 46% similarity (Fabrizio et al. 1997; Brenner and Guthrie 2005)) to elongation factor EF-2 (EF-G in bacteria), a GTPase that catalyses the rearrangement of the ribosome after peptide-bond formation (Tourigny et al. 2013). A fold resembling that of EF-2 was also observed for Snu114 in the early spliceosomal structures solved by cryoEM (Nguyen et al. 2015). Snu114 possesses five GTPase motifs in domain I (**Figure 1-3 A**). The G''-domain is an insertion unique to Snu114 and EF-2 GTPases. In contrast to most GTPases, both Snu114 (Fabrizio et al. 1997) and EF-2 possess a histidine instead of the usual catalytic glutamine in switch II. In contrast to EF-2, Snu114 has an extended N-terminus. Normally, GTPases are assisted by GTPase activating proteins (GAP); interaction with these factors enables the placement of a water molecule into the active loop in switch II, close to the gamma-phosphate of the guanosine triphosphate (GTP) molecule to facilitate catalysis (Vetter and Wittinghofer 2001; Wittinghofer and Vetter 2011). After GTP hydrolysis, the GDP is generally exchanged for GTP by another protein called GTP exchange factor (GEF). The GAP of EF-2 is thought to be the rotated-state of the ribosome but it appears to have no GEF (Lin et al. 2015). Neither GEF nor GAP are known for Snu114.

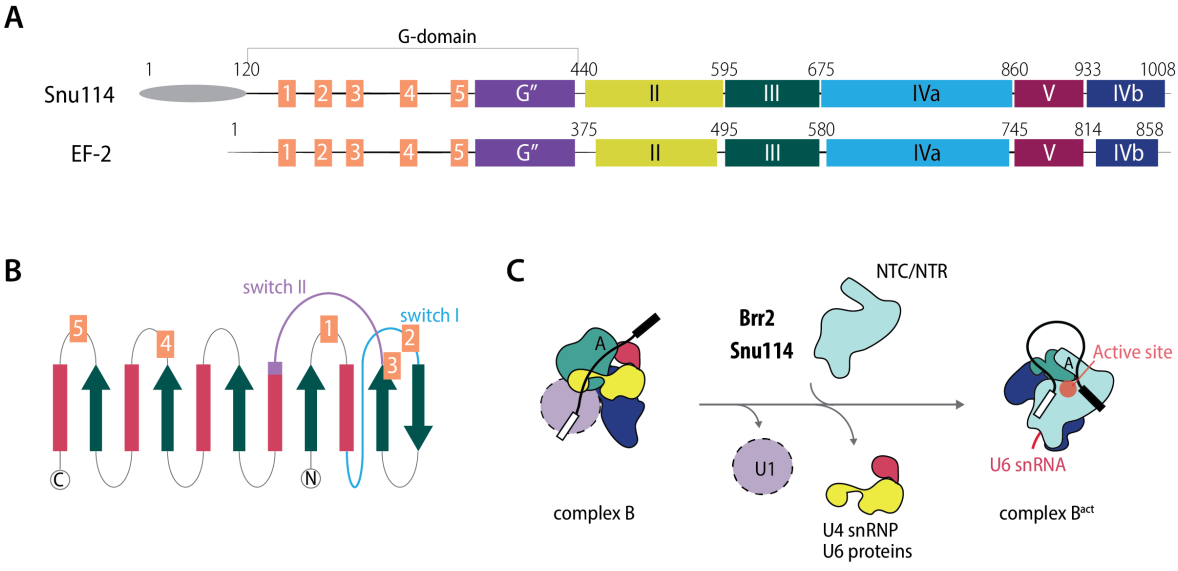


Figure 1-3: The protein Snu114 and its suggested role in spliceosome activation. (**A**) Snu114 has five common domains with EF-2 in addition to its unique N-terminus. Five GTPase motifs locate to the G-domain and are conserved between Snu114 and EF-2. The domain organisation for *S. cerevisiae* Snu114 was adopted from (Brenner and Guthrie 2005) and overlaid to the protein sequence of human EF-2 as in (Fabrizio et al. 1997). (**B**) Topology diagram of a general G-domain showing the location of the five GTPase motifs. The so-called 'switch-regions' are indicated in purple and blue. α-helices are depicted in red and β-strands in green (Wittinghofer and Vetter 2011). (**C**) Snu114 was supposed to give the initial signal for Brr2-mediated unwinding during spliceosome activation. Unwinding of the U4/U6 snRNA duplex leads to the release of U4 snRNP and the U6 proteins, while it allows U6 snRNA to leave its inhibited position and participate in the formation of the active site. The rearrangement also involves the association of the NTC/NTR proteins.

Indeed, Snu114 was found to be essential for splicing. Deletion of the Snu114-specific N-terminal domain causes a temperature sensitive phenotype in yeast and accumulation of stalled spliceosomes (Bartels et al. 2002). A similar effect was seen for mutants with an altered GTP-binding site (Bartels et al. 2003). Snu114 was proposed to play a role in spliceosome activation as the combination of mutations in the N-terminus and in Prp8 or Brr2 lead to a lethal phenotype (Brenner and Guthrie 2005). In a proposed model, Snu114 GTPase transfers an activation signal via Prp8 to the helicase Brr2 (Bartels et al. 2002). As described above, Brr2 catalyses the unwinding of the U4/U6 duplex to allow formation of the catalytic centre during the complex B to B^{act} transition (Brr2 helicase function is described in detail in **1.3.3**). Prp8, likewise an integral part of U5 snRNP, is the largest spliceosomal protein forming the active site cavity and indeed contacts Snu114 (Grainger et al. 2009; Galej et al. 2013; Nguyen et al. 2015). On the other hand, Prp8 interacts with Brr2 and its C-terminus was shown to have an inhibitory function for Brr2 unwinding (Mozaffari-Jovin et al. 2013).

The many similarities between Snu114 and EF-2 led to the speculation of Snu114 acting as the activating GTPase in splicing, for which the correctly assembled complex B could serve as a GAP in a similar way as the rotated state of the ribosome activates EF-2 for GTP hydrolysis.

Two independent studies tried to understand the role of Snu114 in a more mechanistic way. In the first one, Snu114 was claimed to activate Brr2 helicase activity through GTP hydrolysis (Bartels et al. 2002; Bartels et al. 2003) acting like its homologue EF-2. The second study, however, showed that nucleotide binding and not hydrolysis is the activating signal for Brr2, which implies Snu114 to be similar to a small switch-like GTPase (Small et al. 2006). A third hypothesis is drawn from structural analysis of Snu114-containing spliceosomal complexes, and ascribes a solely structural role for the GTP molecule during splicing (Nguyen et al. 2016).

Whatever the role of Snu114 in spliceosome activation, the way in which the major rearrangement between complex B and B^{act} is initiated and accomplished on a mechanistic level remains elusive.

1.3 Remodelling of spliceosomal complexes by RNA helicases drives the splicing reaction

This introductory section will focus on the spliceosomal helicases, which have multiple roles during the splicing process. The first paragraph (**1.3.1**) will give a general overview about the function and architecture of nucleic acid helicases, before each spliceosomal helicase is discussed in more detail (**1.3.2**, **1.3.3**, **1.3.4**). In general, the directionality of the splicing reaction is coupled to the sequential action of the eight spliceosomal helicases. The last paragraph (**1.3.5**) will discuss splicing fidelity which is enhanced by the proofreading activity of spliceosomal helicases preventing the progression of splicing for defective pre-mRNA substrates.

1.3.1 Helicases are involved in all aspects of nucleic acid biology

Nucleic acid helicases are a group of enzymes that are involved in all areas of nucleic acid metabolism. Their activity is essential for DNA replication, recombination and repair as well as for transcription, pre-mRNA splicing, RNA degradation, ribosome biogenesis and translation (Wu and Spies 2013).

Helicases use the energy derived from NTP hydrolysis for various functions. In the classical view, the enzyme couples NTP hydrolysis to separate double stranded nucleic acids into single strands, meaning it unwinds a helix (Pyle 2008). Others translocate along single or double stranded nucleic acid, often referred to as 'translocases' (Fischer and Lohman 2004). NTP-hydrolysis-coupled translocation often coincides with removal of proteins bound to the target strand ('RNPase activity') (Jankowsky et al. 2001). Some helicases are understood to be NTP-dependent switches, which neither unwind a helix nor move directionally, but couple NTP-hydrolysis to a conformational change to induce downstream effects (Singleton, Dillingham, and Wigley 2007).

Some helicases have a preference for their target, which can be DNA, RNA or even hybrid duplexes, while others can act on any nucleic acid (Pang et al. 2002). The helicase core is generally sequence unspecific, however helicases exist that recognise their substrate through additional domains in a sequence specific manner.

Helicases are defined by their conserved nucleotide hydrolysis core engine, which is composed of short amino acid patterns, so-called helicase motifs (Caruthers and McKay 2002). Based on the conservation and organisation of their motifs, helicases are grouped into families. Enzymes of the SF1 and SF2 superfamilies are usually monomeric and comprise at least seven conserved motifs (explained in **Figure 1-4** for SF2), whereas members of the SF3, SF4, and SF5 superfamilies assemble into ring-shaped hexamers (Singleton, Dillingham, and Wigley 2007).

Here we focus on the SF2 family, to which all spliceosomal helicases belong. These monomeric helicases arrange into helicase domain 1 and 2 (HD1 and HD2), both of which resemble a RecA-like fold, and are connected by a linker region (**Figure 1-4 A**) (Fairman-Williams, Guenther, and Jankowsky 2010).

Motif I and II have similarity to the Walker A and B motifs in other ATPases and are sandwiched between the two domains (**Figure 1-4 B**). Motif I (Walker A) with the consensus sequence AxxGxGKT forms a phosphate binding loop. The residues 'GKT' interact with Mg^{2+} and the phosphate region of the nucleotide; the glycine in the triad is required to keep the loop flexible (Walker et al. 1982). Motif II (Walker B), reading DExx, is directly involved in NTP hydrolysis. While the aspartic acid coordinates Mg^{2+}, the glutamic acid activates the attacking water molecule. Motif VI is involved in nucleotide binding and hydrolysis, too (Hall and Matson 1999). Motif Q can provide specificity for

adenosine triphosphate (ATP) by coordinating the adenine base but is not present in all SF2-type helicases (Tanner 2003). Motifs Q, I, II, III, Va and VI compose the nucleotide binding pocket located on both helicase domains. Domain closure is induced upon nucleotide binding (**Figure 1-4 C**). Motifs Ia, Ib, IV, IVa, and V contact the phosphodiester backbone of the nucleic acid target (Singleton and Wigley 2002). The SF-2-type helicase core domain interacts with a minimum of 5 nt (Ozgur et al., 2015).

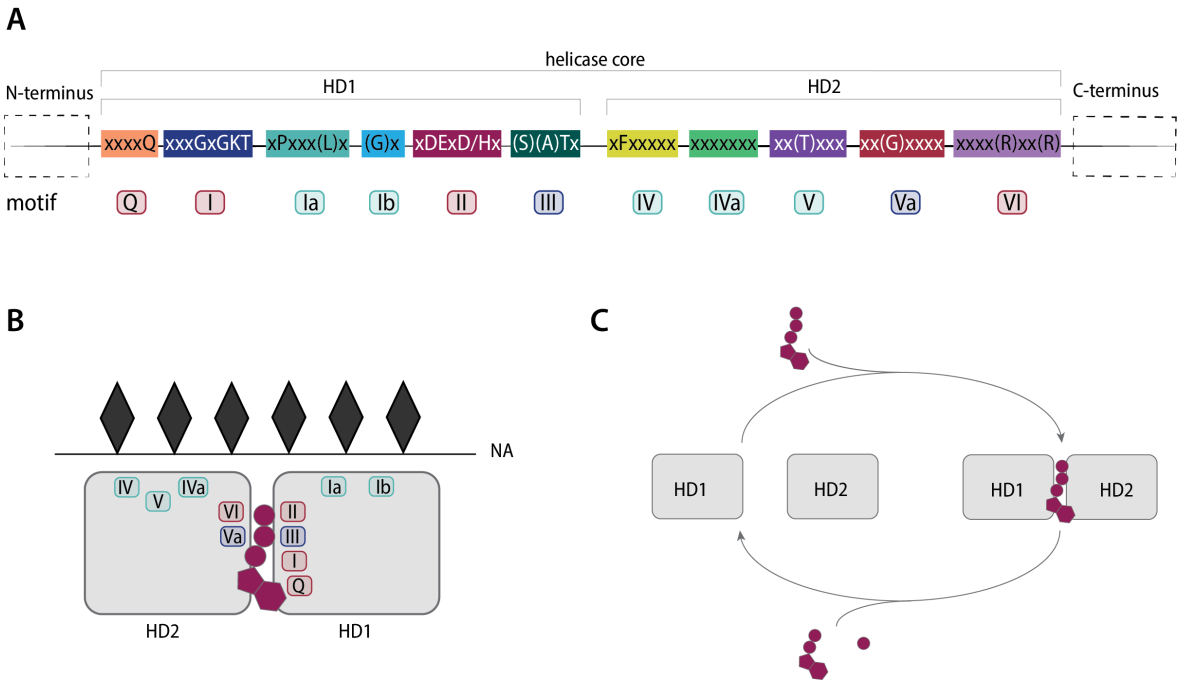


Figure 1-4: The SF2 helicase family. (**A**) Domain organization of the SF2-type helicase core shows the consensus sequence found in helicase motifs. Letters in brackets represent highly but not completely conserved residues. The length of and the distance between the blocks are not depicted to scale. Motifs Q to III are located in helicase domain 1, whereas motifs IV to VI are found in helicase domain 2. Motif IVa is not highly conserved in SF2-type helicases but frequently marked as QxxR. Both C- and N-terminal domains vary between members (Fairman-Williams, Guenther, and Jankowsky 2010). (**B**) Schematic view giving the rough location of the helicase motifs. Both the bound ATP (dark red) and the nucleic acid (NA, black) are indicated. (**A**) and (**B**) show the helicase motifs coloured according to their function. Red: ATP binding and hydrolysis. Blue: Coordination between nucleotide binding and interaction with the nucleic acid. Cyan: nucleic acid binding. (**C**) Upon ATP binding, the helicase domains come into proximity. ATP hydrolysis, generating ADP (adenosine diphosphate) and P_i, causes the helicase domains to swing apart (ATP, ADP and P_i are schematically depicted).

Helicases are often part of large molecular complexes, such as in pre-mRNA splicing where the eight RNA helicases act within the spliceosome. The energy that is required for pre-mRNA splicing is utilised by the spliceosomal helicases to drive the conformational and compositional remodelling of the RNP machinery during the splicing process. RNA duplex unwinding and protein-replacement have been reported as two required helicase functions for splicing (Staley and Guthrie 1998), while translocation on RNA targets and molecular switches were suggested as well. All spliceosomal helicases belong to one of the following sub-classes of SF-2 type helicases:

DEAD-box, Ski2-like, and DEAH-box. Except for the Ski2-like helicase Brr2, which is an integral part of U5 snRNP (Lauber et al. 1996), spliceosomal helicases interact only transiently with the spliceosome.

1.3.2 DEAD-box helicases are involved in spliceosome assembly

DEAD-box helicases derived their name from the single letter code of the four amino acids in motif II. Not all helicases in that family, such as the spliceosomal helicase Sub2, however, read this pattern as D-E-A-D (Zhang and Green 2001). The DEAD-box helicase family is defined by a total of 13 characteristic and conserved sequence motifs, including all helicase motifs of the SF2 superfamily (Linder 2006). Three of the additional motifs are involved in NTP binding and two in the coordination between NTP binding and interaction with the target RNA (**Figure 1-5 A**). DEAD-box helicases make up the largest proportion of the SF2 family and are exclusively RNA helicases. Having the Q-motif (**Figure 1-4 A** and **Figure 1-5 B**), DEAD-box helicases specifically utilise ATP, while most have an enhanced ATPase activity upon RNA binding (Kim et al. 1992; Jacewicz et al. 2014).
Although DEAD-box helicases have been shown to separate RNA duplexes *in vitro* and *in vivo*, they are rather inefficient unwinders (Gilman, Tijerina, and Russell 2017). Instead of translocating on one of the nucleic acid strands for duplex unwinding, DEAD-box helicases directly load onto helical regions and then locally force the strands to open (Cordin et al. 2006). In addition, structured regions of single-stranded RNA can similarly be destabilised upon simultaneous ATP and target RNA binding. Subsequent ATP hydrolysis releases the enzyme from the RNA (Gilman, Tijerina, and Russell 2017). Some DEAD-box helicases have shown to possess RNA annealing activity (Yang and Jankowsky 2005).

Sub2, Prp5 and Prp28, which are the three spliceosomal DEAD-box helicases, share high conversation in their core domains but vary in their C- and N-terminal extensions (**Figure 1-5 A** (Cordin and Beggs 2013)). All three are essential for the assembly phase, indicating they are needed for correct splice site recognition (**Figure 1-5 B**). Most functions for the helicases were deduced by comparing mutants or the absence of the protein to an intact wild-type (WT) system.
In spliceosomal complex E, U1 is the first snRNP to be bound to the pre-mRNA substrate at the 5'SS, while the branchpoint binding protein (BBP) binds to the branch site to bridge the brA and 5'SS by interacting with U1 snRNP specific proteins (Berglund et al. 1997). The BBP itself is thought to recruit the DEAD-box helicase Sub2, whose activity is essential for U2 snRNP to recognise the branch sequence by remodelling the U2 snRNA (Fleckner et al. 1997). Prp5 functions as well in the formation of the pre-spliceosomal complex A and is involved in proofreading the branch site sequence (Liang and Cheng 2015). Its direct RNA target was proposed to be U2 snRNA (O'Day, Dalbadie-McFarland,

and Abelson 1996). Following Prp5 function, U2 snRNP replaces the BBP. Prp5 has long been thought to be an integral part of the spliceosome as it was found during most of the splicing cycle (Kosowski et al. 2009). Later, however, it was shown that tri-snRNP association is blocked until the release of Prp5 (Liang and Cheng 2015) showing that Prp5 is a transient factor. Just after tri-snRNP has joined the spliceosome, the role of Prp28 is to act in the complex pre-B, where it mediates the transfer of 5'SS pairing with U1 to U6 snRNA (Staley and Guthrie 1999), tethering the 5'SS to U6 snRNA for later transfer into the active site of the spliceosome.

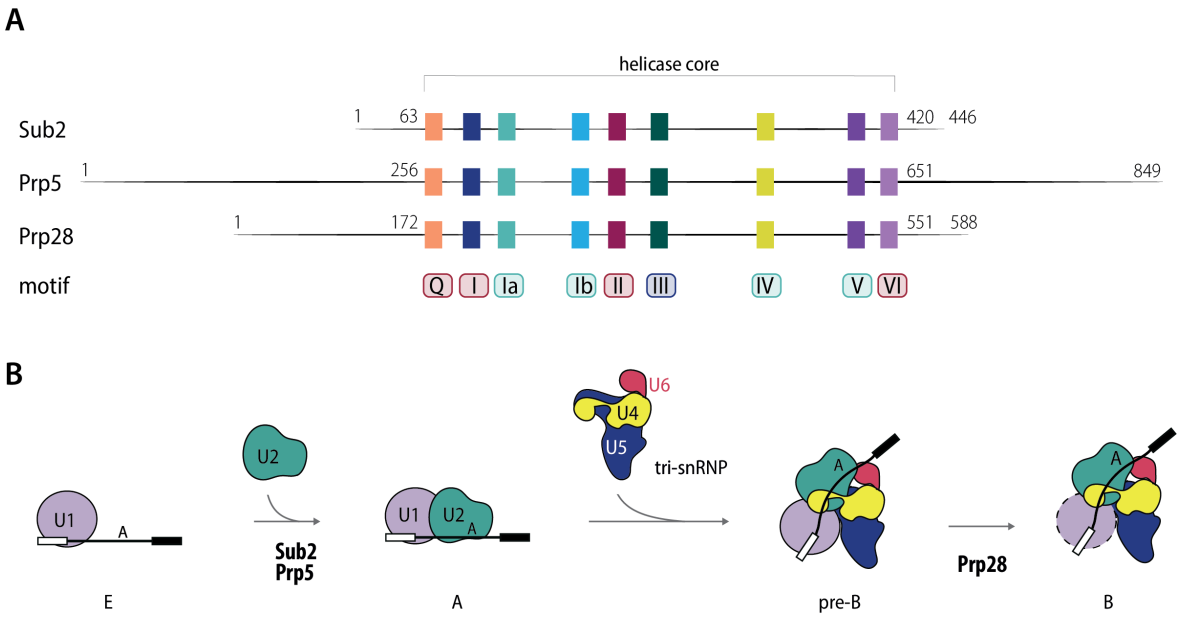


Figure 1-5: Spliceosomal DEAD-box helicases and their role in spliceosome assembly. (A) Sequence alignment of *S. cerevisiae* spliceosomal DEAD-box helicases adapted from (Cordin and Beggs 2013). Nine of the thirteen conserved sequence motifs are depicted. The helicase motifs are coloured according to their function. Red: ATP binding and hydrolysis. Blue: Coordination between nucleotide binding and interaction with the RNA. Cyan: RNA binding. (B) After formation of the spliceosomal complex E, Prp5 mediates a remodelling event which replaces the BBP with U2 snRNP whose RNA component forms base pairs with the BPS. This step requires additional remodelling by Sub2. Prp28 is required for the 5'SS transfer from U1 to U6 snRNA to promote the complex pre-B to B transition.

The function of DEAD-box helicases is essential for normal growth of *S. cerevisiae*. Knocking out some of the spliceosomal components, which are normally destabilised by the acting helicase and which are thought to stabilise certain RNA structures, however, could bypass the function of the corresponding helicase (Kistler and Guthrie 2001). Therefore, a rather passive role was ascribed to spliceosomal DEAD-box helicases (Cordin and Beggs 2013) in agreement with an ATP-induced molecular switch without active RNPase or unwinding activity.

How the three helicases act in molecular detail on their target RNA, and how their activity is regulated in the context of the spliceosome remain unclear.

1.3.3 The Ski2-like helicase Brr2 is essential for spliceosome activation

The helicase Brr2 belongs to the Ski2-like ATPase subfamily of SF2 helicases. Ski2-like helicases utilise exclusively ATP for their function, determined by their possession of the Q-motif. Between helicase motif V and VI, Ski2-like helicases contain a characteristic β-hairpin (Jankowsky 2011). Studies on Ski2-like family members found a correlation with the 3' to 5' unwinding polarity and suggested a function for the hairpin in duplex separation (Buttner, Nehring, and Hopfner 2007). Brr2 is an unusual Ski2-like helicase in possessing two helicase modules of which only the N-terminal module is active (**Figure 1-6 A**); it was shown to interact with RNA *in vivo* without the C-terminal part of the protein and has both ATPase and unwinding activity (Kim and Rossi 1999).

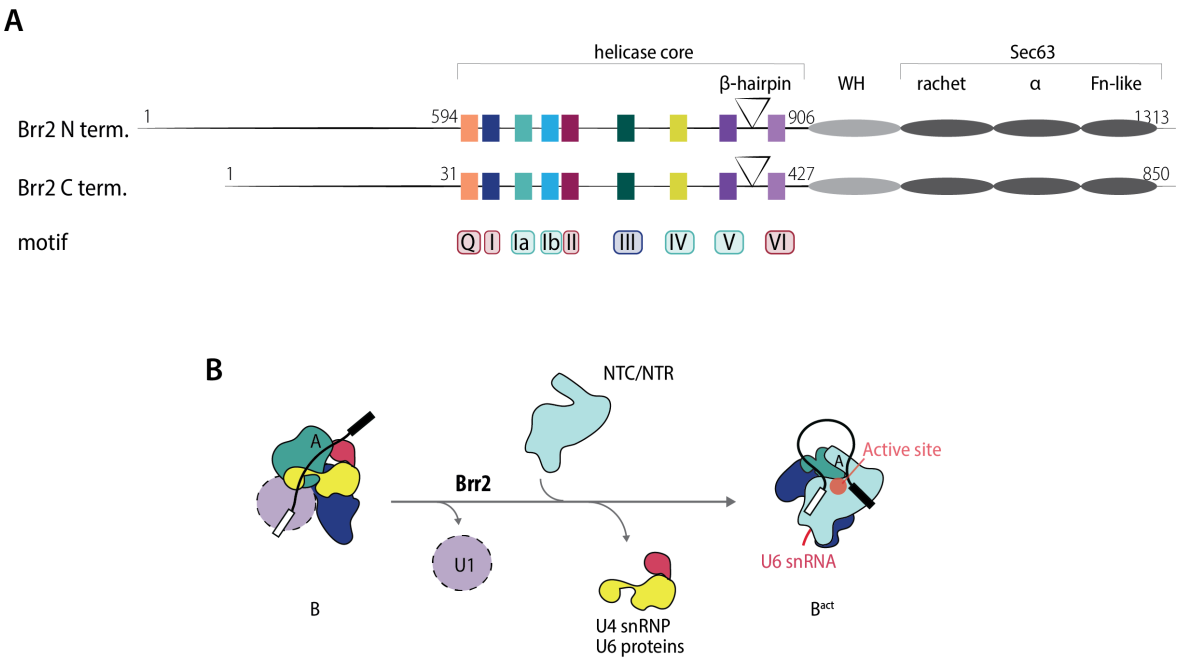


Figure 1-6: The Ski2-like spliceosomal helicase Brr2 and its role in spliceosome activation. (**A**) Sequence alignment of the two helicase cassettes of *S. cerevisiae* Brr2 adapted from (Cordin and Beggs 2013) showing the rough location of the helicase motifs. The motifs are coloured according to their function. Red: ATP binding and hydrolysis. Blue: Coordination between nucleotide binding and interaction with the RNA. Cyan: RNA binding. The triangle between domain V and VI indicates the β-hairpin characteristic for Ski2-like helicases. The C-terminal Sec63 domain is connected through a winged helix (WH) domain and consists of the following subdomains: a ratchet domain, a short α-helical domain (α) and a fibronectin-like domain (Fn-like). (**B**) After formation of the pre-catalytic complex B, spliceosome activation happens through the release of U4 snRNP and U6 proteins, which allows the spliceosomal RNA components to fold into the catalytic core (Madhani and Guthrie 1992; Newman and Norman 1992; Sun and Manley 1995). This rearrangement comes along with the association of the nineteen and nineteen related complexes (NTC, NTR) (Chan et al. 2003) and is initiated by the helicase Brr2, which unwinds the U4/U6 snRNA duplex (Raghunathan and Guthrie 1998; Laggerbauer, Achsel, and Luhrmann 1998). The time point when U1 snRNP is completely released from the spliceosome is unclear and might happen during activation or earlier (compare dashed outline).

Brr2 is the only spliceosomal helicase that is an integral part of the spliceosome, belonging to the U5 snRNP specific proteins. It is already pre-loaded on U4 snRNA in the context of tri-snRNP when it gets recruited to the spliceosome (Nguyen et al. 2016). In complex B, Brr2 still locates to the same

position where it is bound by complex B specific proteins (Plaschka, Lin, and Nagai 2017), ready to unwind the U4/U6 snRNA duplex in a 3' to 5' direction (Laggerbauer, Achsel, and Luhrmann 1998; Raghunathan and Guthrie 1998; Santos et al. 2012) to dissociate U4/U6 snRNP proteins and the U4 snRNA (**Figure 1-6 B**). The liberated U6 snRNA, can refold leading to base pairing with U2 snRNA and the formation of the active site of the spliceosome. So far, the mechanism of Brr2-mediated unwinding has mainly been studied in minimal systems using purified material but lacking the natural context of a complete spliceosome (Mozaffari-Jovin et al. 2012).

The architecture of spliceosomal complex B^{act} locates Brr2 at the periphery of the complex (Yan et al. 2016a). Brr2 has changed location in complex C (Galej et al. 2016), while its location could not be determined in the structures of some spliceosomal complexes (Yan et al. 2015; Fica et al. 2017). This raises the question, how stable Brr2 is integrated into the spliceosome.
In yeast two hybrid experiments, Brr2 helicase was shown to interact with the DEAH-box helicases Prp2 and Prp16 (van Nues and Beggs 2001; Chen et al. 2013), and this interaction could be verified for Prp16 shown in the cryoEM structure of complex C (Galej et al. 2016) (spliceosomal DEAH-box helicases are discussed in **1.3.4**). Questions such as the following remain open: Does Brr2 have a function after spliceosome activation, when it interacts with other helicases or is the helicase stored there for later activity? One study, for example, found Brr2 to be involved in in the rearrangement of the spliceosome towards a second step conformation (Hahn et al. 2012) in line with the Brr2-Prp16 interaction. In addition to its genuine role in unwinding the U4/U6 snRNA duplex during activation, Brr2 was proposed to be involved in disassembly of the ILS together with the DEAH-box helicase Prp43 (Small et al. 2006; Fourmann et al. 2013).

1.3.4 DEAH-box helicases remodel the spliceosome during catalysis and disassembly

Helicases of the DEAH-box family fold into a structure similar to that of the DEAD-box helicases. Their barely conserved N-terminal domain is thought to be involved in target recognition and recruitment (Fairman-Williams, Guenther, and Jankowsky 2010). The C-terminus is well conserved and consists of structured domains (**Figure 1-7 A**), which stabilise the helicase core (He, Andersen, and Nielsen 2010). In contrast to DEAD- and Ski2-like helicases, the Q-motif coordinating the adenine nucleobase is absent in DEAH-box helicases, making them unspecific for the energy-delivering NTP (Tanaka and Schwer 2005). Similar to Brr2, DEAH-box helicases translocate along the RNA strand in order to unwind duplex regions with a preferred 3' to 5' directionality. *In vitro* experiments have shown that DEAH-box helicases cannot load onto structured RNA regions directly, but require a single stranded overhang for initial binding (Tanaka and Schwer 2005). In addition, opening of the two helicase domains was shown to be required for proper loading.

It is thought that a pair of extended β-hairpin motifs, unique to DEAH-box helicases, trap a stretch of the RNA between each other and operate the translocation by incorporating an additional base on the 5'-end while excluding one on the 3'-end (He et al. 2017). During this directional translocation, the complementary strand is displaced (Pyle 2008).

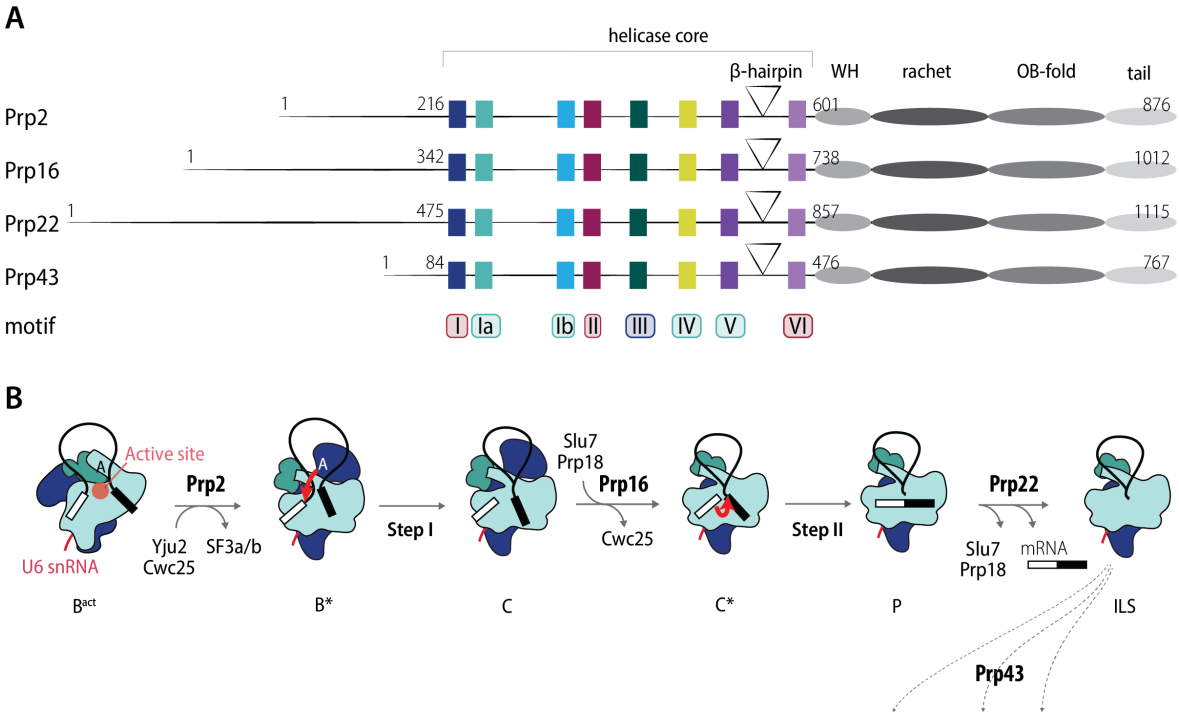


Figure 1-7: Spliceosomal DEAH-box helicases and their role during the catalytic stages and disassembly. (**A**) Sequence alignment of *S. cerevisiae* Prp2, Prp16, Prp22, and Prp43 adapted from (Cordin and Beggs 2013) showing the rough location of the helicase motifs. The motifs are coloured according to their function. Red: NTP binding and hydrolysis. Blue: Coordination between nucleotide binding and interaction with the RNA. Cyan: RNA binding. The triangle between domain V and VI indicates the 5'-β-hairpin insertions, while the 3'-β-hairpin was proposed to be motif Ib (He et al. 2017). The C-terminal domain is composed of a winged-helix (WH), a ratchet domain, and an OB-like fold. (**B**) Schematic view for DEAH-box helicase action during splicing. Prp2 initiates a remodelling event that brings the 5'SS and the brA into proximity to allow the first step of splicing. Binding sites for the branching factors Cwc25 and Yju2 are generated, while SF3a/b complexes, which are part of U2 snRNP, are released. After the reaction spliceosomal complex C is formed, containing the cleaved 5'-exon and the intron-lariat-3'-exon intermediate. Prp16 joins the complex, promoting release of the branching factors and binding of the exon ligation factors Slu7 and Prp18, which in turn leads to docking of the 3'SS into the active site of the spliceosome (complex C*). After the second step of splicing, yielding complex P, Prp22 mediates the release of the ligated exons. Subsequently, the intron-lariat-spliceosome (ILS) is disassembled into free U2, U5, and U6 snRNPs by Prp43 (and Brr2) for recycling, while the excised intron-lariat is released. Helicases are depicted at the position where they are proposed to have their main function. Neither the stage when they join nor when they leave the complex is indicated.

Four DEAH-box helicases are involved in pre-mRNA splicing: Prp2, Prp16, Prp22, and Prp43 (**Figure 1-7**). The helicase Prp2 is involved in remodelling the spliceosome prior to the first catalytic step (Kim and Lin 1996). It acts on spliceosomal complex B^{act} and promotes juxta positioning of the 5'SS and the branch point adenosine for the branching reaction (Warkocki et al. 2009). In addition, Prp2's action is also needed to promote binding of the branching factors Yju2 and Cwc25 (Chiu et al. 2009;

Krishnan et al. 2013). The CryoEM structure of complex B^{act} shows that the branch helix is held by the SF3b complex (Yan et al. 2016a). If Prp2 translocates from its binding position downstream of the branch helix towards the brA on the intron, the SF3a and SF3b complex (part of U2 snRNP) would likely dissociate, which in turn would allow the branch helix to move into the active site of the spliceosome (Lardelli et al. 2010). Unfortunately, the resolution around Prp2 was not high enough in the complex B^{act} reconstruction to model the complete RNA path of the intron (Yan et al. 2016a).

After branching, the helicase Prp16 is required to reposition the substrate in a way that allows the 3'SS to reach the active site. In addition, Prp16 remodelling also causes branching factors to be released. At the same time, binding sites for exon ligation factors Slu7 and Prp18 are generated which allows the complete transition towards catalytically active complex C* (Ohrt et al. 2013).

After Prp16 is released from the spliceosome, the helicase Prp22 binds at a similar location on the particle (Galej et al. 2016; Fica et al. 2017). The exons will be ligated during the second catalytic step and Prp22 is the helicase required to release the spliced RNA from the spliceosome (Company, Arenas, and Abelson 1991). In parallel, exon ligation factors are released. Prp22 was demonstrated to unwind duplex RNA *in vitro* (Schwer and Gross 1998). The helicase was proposed to open contacts between the spliced RNA formed with spliceosomal proteins and U5 snRNA by translocating along the second exon in a 3' to 5' direction, and thereby releasing the ligated exons from the spliceosome (Schwer 2008). In additional studies, Prp22 was found to bind on the intron close to the 3'SS before exon ligation (McPheeters, Schwer, and Muhlenkamp 2000).

These two binding locations for Prp22 imply repositioning of the helicase on the RNA substrate, however, the mechanism remains elusive.

Prp43 plays a major role in disassembly of the resulting intron-lariat spliceosome (ILS) (Mayas et al. 2010). As described above, Brr2 was found to be involved in the disassembly process, too (Small et al. 2006). Despite both Prp43 and Brr2 being extensive targets for structural analysis (He et al. 2017; Tauchert et al. 2017), their precise RNA target and their function in the context of spliceosome disassembly, however, remain quite unknown (Fourmann et al. 2013). Prp43, for example, was proposed to disrupt base pairing between U2 snRNP and the branch sequence, which would lead to the release of the three snRNPs and the excised intron-lariat (Fourmann et al. 2016), but the helicase was also suggested to act on U6 snRNA to disrupt protein-RNA and RNA-RNA interactions (Wan et al. 2017).

Taken together, the biochemical properties of the helicases and most *in vitro* experiments propose a model in which DEAH-box helicases actively unwind RNA duplexes. However, RNA-RNA interac-

tions targeted by spliceosomal helicases are buried within the large body of the spliceosome, making them inaccessible for the transiently, surface bound helicases. Recent experiments with both Prp16 and Prp22 suggest a novel and exciting model, in which the DEAH-box helicase acts from a distance and induces an unwinding effect without actively translocating through the structured RNA parts (Semlow et al. 2016). One of the key observations was that the substitution of RNA by DNA nucleotides, which are located close to the remodelled regions but apart from the initial helicase loading region, did not interfere with helicase activity, despite spliceosomal DEAH-box helicase activity is only stimulated by RNA not DNA (Schwer and Guthrie 1991; Semlow et al. 2016). The result led to the hypothesis that as soon as the helicase is physically stopped from translocating by an obstacle like the spliceosome body, the translocation movement is transferred into a pulling action, which disrupts RNA structures located in the usual translocation direction in a process which the authors named 'winching' (Semlow et al. 2016).

This model remains to be tested to completely unravel the molecular mechanism of 'winching' within the spliceosome and to determine whether this model is universal for all spliceosomal DEAH-box helicases as well as DEAH-box helicases involved in other biological systems.

1.3.5 Spliceosomal DEAD/H-box helicases have proofreading activity to increase splicing fidelity

In general, the directionality of the splicing reaction is coupled to the sequential action of these eight spliceosomal helicases. However, promotion of the splicing reaction is not always favourable. Errors in spliced products, which can potentially lead to dysfunctional or toxic protein products, have to be prevented. To ensure error-free splicing, the spliceosome needs to master the challenge of recognising the minimal sequence elements at the 5'SS, BPS, and 3'SS with high accuracy. Both steps of the splicing reaction are reversible themselves, offering the potential to reject suboptimal pre-mRNA substrates from completing splicing and potentially allow subsequent recognition rounds for sequence elements that had difficulties pairing appropriately with the RNA components of the spliceosome (Tseng and Cheng 2008).

Another mechanism to increase splicing fidelity is mediated during the splicing process by the DEAD/H-box helicases, which were proposed to promote the backwards reaction in case a suboptimal splice site was recognised. Several of the spliceosomal DEAD/H-box helicases have been reported to possess this proofreading activity.

A first indication for proofreading arose from genetic studies, in which a mutation in Prp16 allowed splicing of pre-mRNA substrates carrying branch point mutations and this implied that Prp16 is additionally involved in the discard of abnormal splicing products (Burgess and Guthrie 1993). As discussed above, Prp16's helicase activity normally promotes the rearrangement towards the second

step conformation, which is essential for exon ligation of optimal pre-mRNA substrates (Schwer and Guthrie 1992a). Thus, Prp16 was implicated in proofreading the branching step by promoting a faster splicing process for optimal substrates, while rejecting suboptimal substrates from exon ligation; both of which in an ATP-dependent manner (Koodathingal et al. 2010). Subsequent to the rejection by Prp16, the spliceosome might be disassembled by Prp43 in a discard pathway, which was deduced from an increased frequency of splicing errors in the absence of Prp43 (Mayas et al. 2010). In another study, Prp16 was proposed to shift the equilibrium between competing spliceosomal conformations towards optimal and away from suboptimal substrates (Query and Konarska 2004).

Taken together, two proofreading models were proposed: (1) A kinetic mechanism – The fidelity of the splicing process is enhanced by the Prp16 NTPase activity that acts similar to a timer. Hydrolysis promotes either the productive (exon-ligation) pathway for correct substrates or the discard pathway for suboptimal substrates (Semlow and Staley 2012). (2) A thermodynamic mechanism – Splicing fidelity is enhanced through the stability of competing spliceosome conformations (Semlow and Staley 2012).

In a similar fashion, Prp5, Prp28, Prp2, and Prp22 have been proposed to increase splicing fidelity through proofreading. The DEAD-box helicase Prp5 was indicated in branch site selection, meaning proofreading of the interaction between U2 snRNA and the BPS of the pre-mRNA substrate during spliceosome assembly (Xu and Query 2007). In this study, proofreading activity was determined to be proportional to the ATPase activity of the helicase, supporting the kinetic model. However an alternative ATP-independent role for Prp5 was proposed independently by (Liang and Cheng 2015).

The DEAD-box helicase Prp28 was proposed to proofread selection of the 5’SS on the basis of a genetic screen. Proofreading was shown to be dependent on the 5'SS:U6 snRNA duplex stability (Yang et al. 2013).

The DEAH-box helicase Prp2 was found to proofread the catalytic RNA core of the spliceosome (Wlodaver and Staley 2014). By destabilising the RNA active site, Prp2 might promote proper repositioning of the branch helix.

The DEAH-box helicase Prp22 is proposed to proofread 3’SS selection prior to exon ligation (Schwer and Gross 1998; Mayas, Maita, and Staley 2006) and to lead suboptimal substrates towards the Prp43-mediated discard pathway (Semlow et al. 2016).

In summary, all the above-mentioned studies propose a proofreading mechanism at various steps along the splicing pathway to increase fidelity. The precise mechanism in how each helicase recognises suboptimal substrates and how the reaction is promoted backwards or into a discard pathway in the event of an abnormal substrate remains unknown.

Overall many open questions remain regarding the working mechanism of spliceosomal helicases, in particular: How and when is target recognition and loading onto the RNA initiated? How is the order of subsequently acting helicases maintained and regulated? Do auxiliary factors exist, which regulate the helicase activity? Which parts of the remodelling events are actively promoted by spliceosomal helicases and which are simply a consequence of loosened RNA or RNA-protein interactions?

1.4 Stalling the splicing reaction at specific stages is an indispensable tool for spliceosome research

To study spliceosome biology, no complete *in vitro* system is available, however, spliceosomes can be assembled on synthetic pre-mRNA from yeast whole cell extract *in vitro* (Lin et al. 1985; Newman et al. 1985). With this system in hand and the possibility to stall the splicing reaction at various steps, step specific spliceosomes could be generated. This was essential to gain many of the biochemical and structural insights into the spliceosome and builds the foundation for studying the mechanism of spliceosomal helicases in their natural context.

There are different ways to interfere with the splicing reaction in order to obtain a stall. First, the general concept is listed, before some examples are explained in detail.

(1) The pre-mRNA substrate can be modified in a way to inhibit one of the catalytic steps (Newman et al. 1985; Vijayraghavan et al. 1986; Moore and Sharp 1992).

(2) DEAD/H-box helicase mutants that act in a dominant negative (DN) manner and therefore fail in promoting the transition towards the next spliceosomal conformation can be used to halt the reaction (Plumpton, McGarvey, and Beggs 1994; Schneider, Hotz, and Schwer 2002). When temperature sensitive-mutants exist, yeast cells can be grown under the permissive temperature as usual, while the *in vitro* splicing reaction will halt due to the defective helicase when performed under the restrictive temperature (Schwer and Guthrie 1992b; Kim and Lin 1996). Removal of the helicase through e.g. antibody-assisted depletion or heat inactivation of temperature sensitive mutants has a similar effect.

(3) ATP or Mg^{2+} which are essential for helicase activity, will halt the reaction when used in limiting amounts (Fabrizio et al. 2009).

Finally, the blocked splicing reaction will lead to the accumulation of spliceosomes at the stages prior to the stall. Purification via step specific factors can separate those further. The following paragraph gives examples of stalling methods used at various points in the splicing process.

The activity of the helicase Prp28, for instance, is needed to remodel the spliceosomal complex pre-B into B. Based on the RNA composition, the two complexes are equal, both bind pre-mRNA, which excludes a stall at the complex pre-B stage via the splicing substrate. However, a dominant negative mutant of Prp28, in which the DEAD-motif is altered to AAAD leading to impaired ATP binding and hydrolysis, was reported to stall splicing at the complex pre-B stage (Boesler et al. 2015). In general, a dominant negative helicase mutant is characterised by its ability to strongly associate with the spliceosome in a way that prohibits the WT protein from associating with the complex. To ensure normal growth, which includes functional splicing, yeast cells usually require the WT copy of the protein. Sometimes, the dominant negative helicase can be co-expressed, otherwise it can be added to the splicing reaction from a recombinant source and will then block the function of the WT helicase.

Limiting the ATP concentration can halt the reaction at the complex B stage by impeding Brr2 helicase activity. A minimal but requisite amount of ATP will still allow assembly of the spliceosome, as long as the upstream acting DEAD-box helicases are not impaired by the low ATP level. This strategy was used for instance to determine the cryoEM structure of the pre-catalytic complex B (Chan and Cheng 2005; Plaschka, Lin, and Nagai 2017).

Stalling splicing with a temperature-sensitive mutant of Prp2 (*PRP2-1*) leads to accumulation of spliceosomal complex B^{act}. Temperature-sensitive mutant strains usually grow normally at the permissive temperatures but show a growth defect at higher temperature, indicating the mutated helicase protein is less stable compared to the WT. This property can be used to deplete Prp2-1 from splicing extract. Heat inactivation leads to a defective protein and the helicase cannot associate with the spliceosome but spliceosomal complex B^{act} ΔPrp2 will assemble (Kim and Lin 1996). After the splicing reaction, a dominant negative version of the protein such as Prp2-S378L can be added (Plumpton, McGarvey, and Beggs 1994). The mutation in the SAT residues of motif III (altered to LAT) impairs coupling of NTP-hydrolysis to helicase activity, so that the splicing reaction cannot proceed further.

Acting on spliceosomal complex C, the helicase Prp16 is involved directly after branching to remodel the spliceosome towards a conformation in which the 3'SS is placed into the active site for the second reaction. One way to stall splicing at the complex C stage utilises a dominant negative version of Prp16 such as Prp16-G378A whose alteration of the GKT residues in motif I leads to an NTP binding or hydrolysis defect. Secondly, the RNA substrate can be modified to prohibit the second step of splicing. Substitution of the conserved AG at the 3'SS with e.g. AC impairs with the second catalytic reaction. Similarly, a pre-mRNA substrate that terminates just before the AG-dinucleotide will not proceed further than the complex C stage until an AG-3'-exon is supplied *in trans* (Bessonov

et al. 2008). Depletion of Prp16 from the splicing extract, by e.g. antibody binding and removal, will lead to ΔPrp16 spliceosomes that stall at the same stage (Semlow et al. 2016).

Spliceosomal complex C* represents a stage before exon ligation but after Prp16 remodelling. Equally 3'SS alteration can lead to the C* conformation. For structure determination of the yeast complex, the ribose backbone of the conserved G at the 3'SS was modified to have a 2'-H group instead of the natural 2'-OH group. An additional incubation step with ATP stimulated the remodelling from the complex C towards the C* stage (Fica et al. 2017).

Several dominant negative versions of the helicase Prp22 were discovered, which block mRNA release and thereby stall the splicing reaction at the complex P stage (Tanaka and Schwer 2005). Both Prp22-K512A and Prp22-S635A with an alteration in the GKT residues of motif I minimising ATP hydrolysis and in the SAT residues of motif III which uncouples ATPase from helicase activity, respectively, were used to purify complex P for structural analysis (Liu et al. 2017; Wilkinson et al. 2017).

1.5 Questions to be addressed in this project

As described above, the spliceosome is an extremely dynamic particle. Although major milestones were achieved in recent years determining both the composition and architecture of the spliceosome at various stages throughout the splicing process, much is still unknown about how transitions between the snapshots are enforced and regulated.

This thesis aims at providing a platform to increase our molecular and mechanistic understanding of spliceosome remodelling in the future. For all experiments in this work, the model organism budding yeast (*S. cerevisiae*) was used.

1.5.1 Structural and biochemical analysis of Snu114

The first part of the thesis (**chapter 2**) was initiated to increase the understanding of the role of Snu114 GTPase in spliceosome activation and was aimed to answer the following questions:

(1) What is the overall structure of Snu114 and does the structure imply a function similar to EF-2?

(2) At which stage does Snu114 hydrolyse GTP during the splicing process, and how is GTP hydrolysis triggered?

(3) How does Snu114 influence U4/U6 snRNA unwinding by Brr2?

1.5.2 Analysis of RNA-helicase interactions during spliceosome remodelling

The second part spanning three results chapters focuses on improving the understanding of spliceosomal remodelling event that are mediated by RNA helicases.

Along with an increased interest for structural studies on the spliceosome, stalling and purification approaches leading to compositionally (and structurally) defined complexes moved into the focus of spliceosome research. Simultaneously, next-generation sequencing (NGS) based approaches emerged that enable one to precisely localise an RNA binding protein (RBP) on its target RNA. Combining both offers the opportunity to analyse helicases within a fully assembled RNP and will in turn greatly improve our understanding of their function. Recently, the ambition to address mechanistic questions with NGS approaches rose in the pre-mRNA splicing community (Burke et al. 2018; Chen et al. 2018; Briese et al. 2018), however, an *in vitro* system based on NGS and fully assembled spliceosomes will bridge the gap between structural snapshots and genome wide data of the splicing mechanism.

To address the following objectives, I have conducted experiments in both the laboratory of Dr Kiyoshi Nagai and Prof Jernej Ule in the MRC LMB and Francis Crick Institute, respectively:

(1) Development of iCLIP (individual nucleotide resolution UV cross-linking and immunoprecipitation; UV: ultraviolet) for *in vitro* purified spliceosomes concentrating on Prp16 in spliceosomal complex C
(2) Adaptation of the method to analyse additional spliceosomal helicases
(3) Modification of the method towards the development of a time-resolved translocation profile of a helicase during remodelling

In the thesis, objective (1) is divided into a method development section, which shows the generation of binding profiles for Prp16 helicase and the control spliceosomal protein SmB (**chapter 3**), while data interpretation and fine-tuning of the method for Prp16 is discussed separately (**chapter 4**). Objectives (2) and (3) give an indication towards the feasibility of investigating additional helicases by *in vitro* spliceosome iCLIP and may ultimately lead to monitor remodelling over time (**chapter 5**).

2 Snu114 is apparently a pseudo GTPase

The present chapter aims at understanding the role of Snu114 GTPase in spliceosome activation. When the project was started, we knew about the significant sequence similarity between Snu114 and EF-2 (EF-G in bacteria) (Fabrizio et al. 1997; Brenner and Guthrie 2005) and we thought that detailed structural information on Snu114 would greatly increase our mechanistic understanding of the GTPase in pre-mRNA splicing.

Around the same time, the structure of tri-snRNP was determined to ~6 Å resolution revealing its overall architecture and also showing how the RNA and proteins of the U5 snRNP are organised within tri-snRNP (**Figure 2-1 A,** here shown in high resolution) (Nguyen et al. 2015). Snu114 interacts with the N-terminal domain of Prp8 (Prp8-N), which is in agreement with yeast two hybrid experiments (Grainger et al. 2009) (**Figure 2-1 B**). The architecture of tri-snRNP further revealed that Snu114, the N-terminal domain of Prp8, and U5 snRNA form a discrete substructure within the U5 snRNP (**Figure 2-1 A**). The structure of tri-snRNP itself, however, provided little insight into the function of Snu114 in spliceosome activation.

Since determining high resolution cryoEM structures was considered as an enormous challenge during that time, I aimed to produce a complex built from Snu114, Prp8-N and U5 snRNA for high resolution structural analysis by crystallography in order to reveal a detailed picture of Snu114 in its natural context, bound to Prp8 and U5 snRNA. Determining the structure of the RNP built from Snu114, Prp8-N, and U5 snRNA in different nucleotide states, such as before and after GTP hydrolysis, was of particular interest. These structures should have helped to understand the molecular mechanism of the GTPase and its effects on the interaction with Prp8.

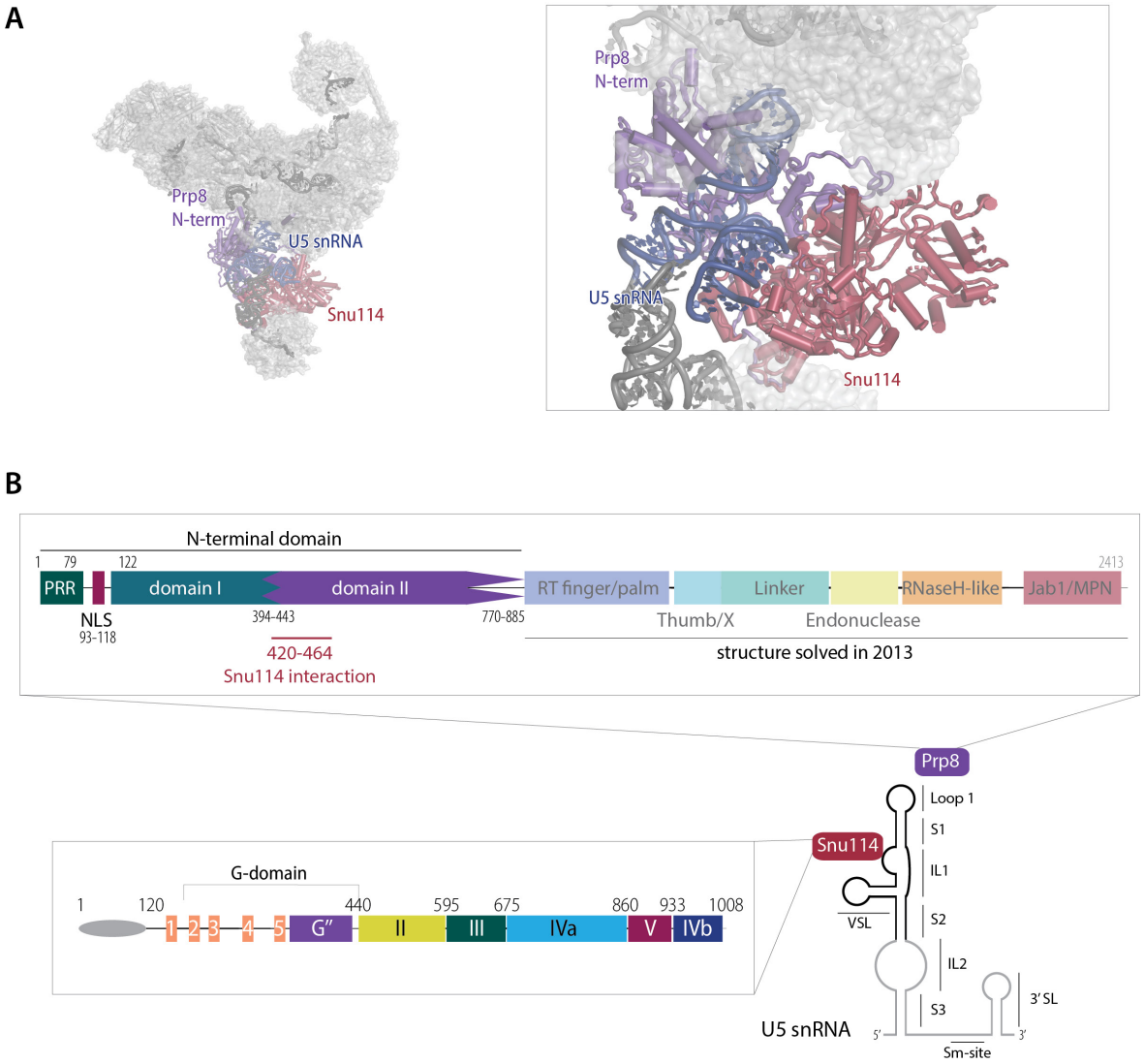


Figure 2-1: Spatial organisation of the Snu114:Prp8-N:U5 snRNA substructure. (**A**) CryoEM structure giving the spatial location of Snu114 (red), the N-terminus of Prp8 (purple) and the 5' stem loop of U5 snRNA (navy) in tri-snRNP as an overview (left) and enlarged (right) (Nguyen et al. 2015); the coordinates used to produce this figure derived from the later, higher resolution structure of (Nguyen et al. 2016). (**B**) Domain organisation of Prp8, Snu114 and U5 snRNA. The N-terminal part of Prp8 is composed of a proline rich region (PRR), a nuclear localisation sequence (NLS), and was predicted to have two structured domains. The three-dimensional structure of the larger part of Prp8, containing RT finger/palm, thumb, linker, endonuclease, RNaseH-like and Jab1/MPN domain, was determined before (Galej et al. 2013). The interaction site with Snu114 locates to the N-terminal domain residues 420-464 (indicated in red). Snu114 possesses five GTPase motifs in domain I. The G"-domain is an insertion unique to Snu114 and EF-2 GTPases. Domains II to V show sequence similarities to other GTPases. The extended N-terminus is unique to Snu114. Snu114 binds to the internal loop 1 (IL 1) whereas Prp8 makes contacts mainly with loop 1 of U5 snRNA (Dix et al. 1998). Grey parts of U5 snRNA were excluded for structural analyses in this work.

2.1 A structural approach to understand the function of Snu114

For the work described in **subchapter 2.1**, I collaborated with C. Oubridge. Experimental planning, conduction, and interpretation were therefore done in collaboration.

2.1.1 Production of the protein complex between Snu114 and Prp8 N-terminal domain

In a first step towards structural investigation, we established a recombinant expression system for the *S. cerevisiae* proteins Snu114 and Prp8-N. Domain organisation in Prp8-N formed the basis for initial construct design (**Figure 2-1**). The approximate boundary between domain I and II as described in (Boon et al. 2006) was analysed by secondary structure prediction, and constructs coding for either of the two domains or spanning both were generated (**Table 8-1** and **Table 8-2** in appendix).

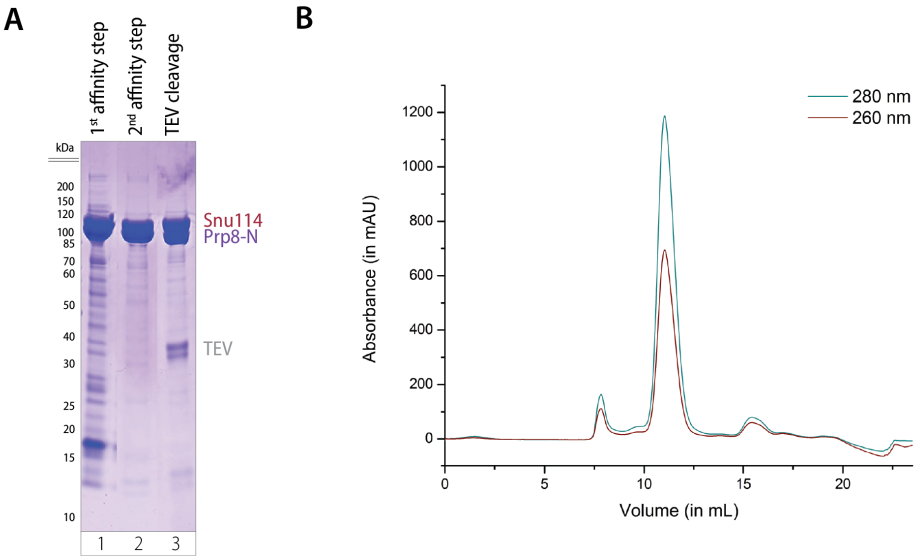


Figure 2-2: Snu114:Prp8-N can be expressed and purified from yeast. (**A**) After calmodulin (lane 1) and nickel nitrilotriacetic acid (Ni-NTA) affinity purification (lane 2) affinity tags are removed by TEV-protease cleavage (lane 3). Snu114, Prp8-N and TEV protease migration locations are indicated on the SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis). (**B**) Gel filtration of the Snu114:Prp8-N using a Superdex 200 column yields a homogenous sample in the major peak. The constructs for this representative Snu114:Prp8-N purification are indicated with a * in **Table 8-1**.

Initially, expression systems in both bacteria and yeast were planned. Despite the suitable length of constructs for *Escherichia coli* (*E. coli*) protein production and testing of various strains and conditions, Prp8-N constructs could not be expressed, or the overall yield remained too low (**Table 8-1** lists the tested constructs). Recombinant expression of Prp8-N domain I also failed in BCY123 yeast cells (**Table 8-1**). Since the interaction between Snu114 and Prp8 was predicted to be strong (Bartels et al. 2002), co-expression of the two proteins was considered to increase solubility and stability, while preventing aggregation. Constructs of Prp8-N including the predicted binding site for Snu114 (Grainger et al. 2009) were finally successfully co-expressed with Snu114 in yeast (**Figure 2-2**), using

the pRS42x co-expression system in which the coding sequence is under the GAL-GAP promoter (MultiSac system). Our constructs were designed for a two-step affinity purification via CBP- and polyhistidine-tags. After optimisation we obtained a clean and stochiometric complex of Snu114:Prp8-N (**Figure 2-2**, lanes 1 and 2) (**7.3.1 Materials and methods**, **Appendix Table 8-1** for successful expression of different constructs). TEV (tobacco etch virus) protease cleavage (Polayes et al. 1998) efficiently removed the CBP-His-tag from the N-terminus of Prp8-N (**Figure 2-2**, lane 3). Subsequent gel filtration for additional purification, removal of the TEV-protease and released affinity tags shows that most of the protein complex elutes as one soluble population (**Figure 2-2 B**, main peak); only a small fraction eluted in the void volume where aggregations are found normally (**Figure 2-2 B**, peak at 7.5 mL).

In total, an overall yield of about 0.15 mg protein complex per 1 L yeast culture could be produced from the yeast expression system, sufficient for reconstitution with U5 snRNA and subsequent crystallisation trials. Co-expression of the two proteins, yielding to a stoichiometric complex, is beneficial for subsequent reconstitution with U5 snRNA.

2.1.2 RNA binding studies show that a complex can be formed from the recombinant protein dimer and *in vitro* transcribed RNA

In the living cell, Prp8, Snu114 and several other proteins assemble with U5 snRNA to form the U5 snRNP, a building block of the spliceosome (Nancollis et al. 2013). As described before, structure determination of Snu114 was attempted in a natural context, meaning bound to Prp8-N and U5 snRNA, in order to learn most about its biological function. Snu114 binds to the internal loop 1 (IL 1) whereas Prp8 makes contacts with loop 1 of U5 snRNA (Dix et al. 1998).

For crystallographic sample preparation, a U5 snRNA fragment should be generated by *in vitro* transcription. The shorter version of U5 snRNA, containing the binding sites for Snu114 and Prp8, was designed based on the architectural information from tri-snRNP with the intention to produce a distinct and compact particle suitable for crystallography. Thus, a truncated U5 snRNA, consisting of only an apical region of the 5' stem-loop but lacking the Sm-binding site was constructed (**Figure 2-1 B** and **Figure 2-3 A** for U5 snRNA A).

In addition, the truncated U5 snRNA A was extended by the four-nucleotide motif CGCG or CATG on the 3'-end (**Figure 2-3 A**, U5 snRNA B and C) to allow helix formation by base-pairing of the overhang between two molecules and thus to lead to a two-fold symmetry of particles later in crystallography. After *in vitro* transcription, the homogeneity of all three RNA constructs was analysed by denaturing polyacrylamide gel electrophoresis (PAGE) and mass spectrometry (**Figure 2-3 B** shows denaturing PAGE).

Reconstitution of the protein dimer with the RNA was achieved by mixing the protein in high salt and the RNA in low salt in such volumes as reached a final concentration of 0.2 M KCl. Electrophoretic mobility shift assays (EMSA), in which the protein concentration was increased, were used to find the optimal protein:RNA ratio (**Figure 2-3 C**). In later reconstitutions, a ratio of 1.25:1 protein:RNA was used, which leads to nearly complete complex formation (**Figure 2-3 C**, lane 6).

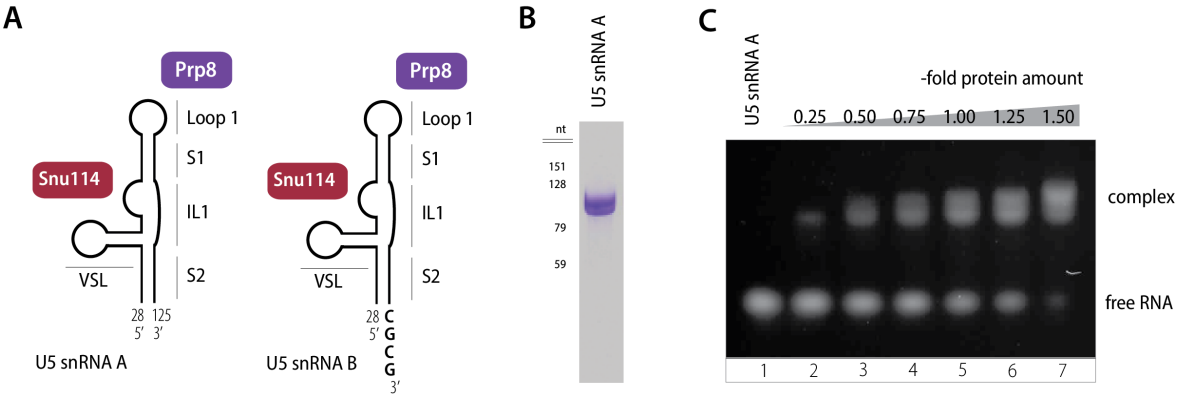


Figure 2-3: U5 snRNA forms a complex with Snu114:Prp8-N. (**A**) U5 snRNA constructs used in this study. U5 snRNA A is formed of nucleotides 28-125 of U5 snRNA. U5 snRNA B is extended with the 4-nucleotide motif CGCG at the 3'-end. U5 snRNA C was similarly extended with the motif CATG (not depicted). (**B**) Migration profile of *in vitro* transcribed U5 snRNA A on a denaturing PAGE (polyacrylamide gel electrophoresis). (**C**) EMSA of U5 snRNA A with increasing protein concentration (Snu114-His_8:Prp8_122-770). The numbers give the molar ratio of protein-dimer compared to 0.6 µM U5 snRNA (final concentration).

Initially, concentrating the complex led to a loss of sample which could be improved by increasing the salt concentration while decreasing the volume. Homogeneity of the complete RNP was tested by MALS (multi-angle light scattering), a method that determines the molecular mass and hydrodynamic radius. A homogenous behaviour of complexes formed from Snu114:Prp8-N with the different U5 snRNA constructs (compare **Figure 2-3 A**) could be verified by MALS (data not shown).

In summary, the protein-RNA complex could be reconstituted and produced in sufficient amounts for crystallisation trials. Consequently, crystal plates covering 1248 conditions for three different constructs (Snu114-His_8:Prp8_122-770:U5 snRNA A, Snu114-His_8:Prp8_1-770:U5 snRNA A, Snu114:Prp8_122-770: U5 snRNA A) were set up. Promising protein-RNA complex-containing crystals were not found. However, the behaviour of the complex in different conditions reveals some trends. Most screens showed precipitate in about half of the wells, whereas the other half contained clear droplets or droplets with phase separation indicating a suitable concentration of the target complex in general. Complexes reconstituted with U5 snRNA B or C could potentially crystallise due to the forced pseudosymmetry but have not been tested to date.

The overall set up for producing the Snu114:Prp8-N:U5 snRNA RNP is quite versatile as the three complex components can easily be altered individually. Thus, in parallel to modifying the RNA construct, we targeted the protein components for optimisation. Limited proteolysis performed on the reconstituted protein-RNA complex was used to identify disordered or unstructured regions in the protein components that could hinder crystal formation. Prp8-N was digested with trypsin, chymotrypsin or subtilisin into stable fragments. Trypsin, for instance, cuts Prp8-N at the positions 122 and 318 (**Figure 2-4 A**). Glu-C and Asp-N did not generate distinct fragments.

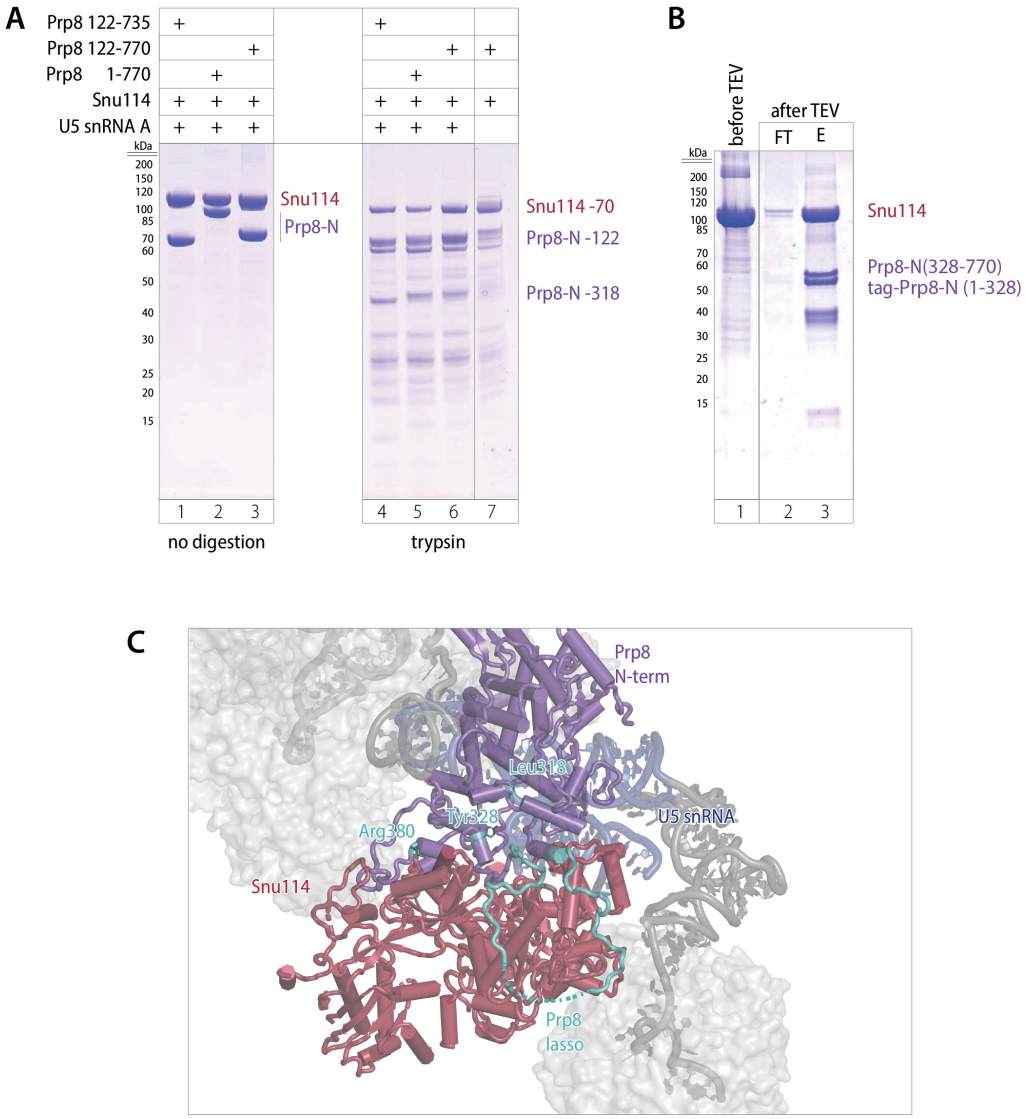


Figure 2-4: Prp8-N cannot be subdivided into two distinct domains. (**A**) SDS-PAGE after limited proteolysis with trypsin of protein and protein:RNA complexes. The used constructs are indicated in the figure. Major cleavage fragments are labelled indicating the aa position, which we interpreted as the cleavage site. (**B**) SDS-PAGE showing unsuccessful separation of cleaved fragments (lanes 2 and 3) after TEV protease cleavage of the protein:RNA complex. (**C**) Close-up view of the location of introduced TEV cleavage sites (labelled Leu318, Tyr328, Arg380) and the lasso region (all light cyan) in tri-snRNP (Nguyen et al. 2016). The lasso region was transferred from the *S. pombe* structure (Yan et al. 2015) by sequence alignment and locates to residues 413–474 in *S. cerevisiae* Prp8. Unmodeled parts in the lasso region are connected with a dashed line.

New constructs were designed based on masses of the proteolytic fragments, taking into account the amino acid preferences of the protease and ensuring the binding site of Snu114 on Prp8 (residues 420-464) was included. To ensure proper formation of the Snu114:Prp8-N dimer, a TEV-protease cleavage sites for trimming after expression were included instead of shortening the expression construct directly. For Snu114, a construct was designed placing a TEV-protease cleavage site near its N-terminus. Three constructs with TEV-protease cleavage sites at position 318, 328 and 380 were designed for Prp8; all three sites locate to the predicted boundary between domain I and II. All constructs could be expressed; however, the protease sites at positions 328 and 380 in Prp8 became accessible to TEV protease only after insertion of a spacer sequence upstream (**Figure 2-4 B**, lane 3 for cleavage at position 328). Reconstitution with RNA was still successful after TEV protease cleavage. The N-terminal region of Prp8 to be cleaved by TEV protease includes the polyhistidine-tag, hence the fragment should be separated from the protein dimer in a subsequent nickel nitrilotriacetic acid (Ni-NTA) affinity step. Despite incubation in chaotropic and denaturing conditions or high salt (high concentrations of KSCN, 2 M urea, 2 M KCl) prior to Ni-NTA affinity purification, the cleaved fragment stayed tightly bound to the rest of the complex (**Figure 2-4 B**, lanes 2 and 3).

Thus, we interpreted that the tight binding of the cleaved fragment must have an important role in maintaining the Prp8 structure itself or the interactions with Snu114.

2.1.3 Concluding remarks for the crystallisation attempts of Snu114:Prp8-N:U5 snRNA

Before this project yielded any structural results, near atomic-resolution information on the Snu114:Prp8-N:U5 snRNA subcomplex became available in the context of the intron-lariat spliceosome, whose structure was obtained by cryoEM (Hang et al. 2015; Yan et al. 2015). It became clear why Snu114 and Prp8 could be only expressed together (**Figure 2-2**) and why the cleaved Prp8-N fragments were impossible to separate (**Figure 2-4 B**). An extended unstructured loop of Prp8, named lasso for the *S. pombe* protein equivalent (Yan et al. 2015), emerges from the body of the protein and wraps around Snu114 (**Figure 2-4 C**). The domain boundary proposed by (Boon et al. 2006) indeed identifies unstructured protein, namely the lasso region. The flanking regions, however, do not organise into two distinct domains but form one rigid body. By lying on the surface of the protein it becomes clear that the introduced TEV-protease cleavage sites were accessible during limited proteolysis without dividing the protein into subdomains.

At about the same time when the structure of intron-lariat spliceosome from *Saccharomyces pombe* was published, the resolution of tri-snRNP improved until molecular interactions within the relevant region of U5 snRNP could be observed (Nguyen et al. 2016). Based on the density in tri-snRNP, the authors concluded that Snu114 binds a GTP molecule that is firmly coordinated by side

chains of the GTPase (**Figure 2-5**). The authors claimed that a different nucleotide state would not fit well into the conformation of the binding pocket, while also suggesting that Snu114 must rather accommodate a GTP than a GDP molecule in the intron-lariat spliceosome structure.

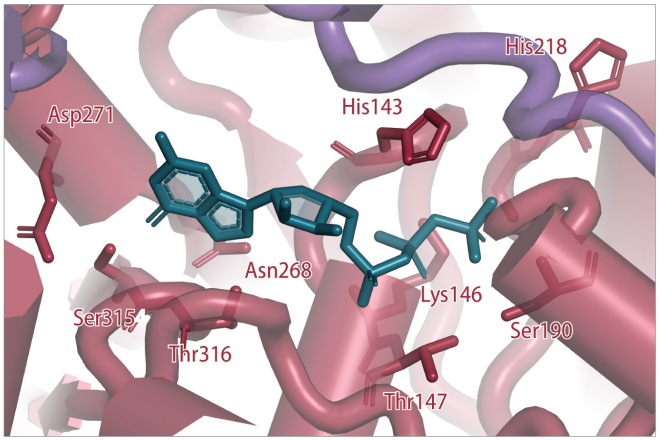


Figure 2-5: Snu114 binds to GTP in tri-snRNP. Zoom into the GTP-binding pocket of Snu114. Snu114 is shown in red, Prp8 in purple and the GTP molecule in cyan. Residues which were described to coordinate the GTP (Nguyen et al. 2016) are labelled with their residue number and shown as sticks. His218 is indicated in the figure since the equivalent residue has catalytic activity in EF-2.

Although both particles reveal the structure of the subcomplex formed by Snu114:Prp8-N and U5 snRNA and although the nucleotide state of Snu114 GTPase is structurally determined in both static spliceosomal complexes, neither of them can explain the role of Snu114 for spliceosome activation. In addition, there is no indication if GTP is exchanged or hydrolysed to GDP at a different stage during pre-mRNA splicing.

2.2 Snu114 binds GTP throughout the splicing cycle

The lack of explanation for the role of Snu114 in activation inspired us to initiate genetic and biochemical analysis of the GTPase. The following two open questions were aimed to be addressed: (1) Does Snu114 hydrolyse GTP during splicing and, if so, at which stage does this happen? (2) How does Snu114 influence U4/U6 snRNA unwinding by Brr2?

As mentioned in the **introductory subsection 1.2**, two independent studies aimed at understanding the role of Snu114 through biochemical experiments (Bartels et al. 2003; Small et al. 2006). For their analysis, both studies made use of the mutant Snu114-D271N, in which the nucleotide binding specificity is altered for xanthine instead of guanine nucleotides. This modification is commonly used to conduct GTPase experiments in an orthogonal system. Working with xanthine triphosphate (XTP), which does not normally occur naturally, allows one to exclude effects derived from endogenous GTP in the system (Bishop et al. 2000).

(Bartels et al. 2002; Bartels et al. 2003) claim that Snu114 activates the helicase activity of Brr2, and this requires GTP hydrolysis. The experiments which led to their conclusion were set up with early spliceosomes, captured under limiting ATP concentrations which causes splicing to stall at the complex B stage. After incubation with different nucleotides, the amount of released U4 snRNA was determined as a measure for spliceosome activation. The highest amount of U4 snRNA release could be obtained with a mixture of hydrolysable XTP/GTP and ATP. Non-hydrolysable XTP/GTP, however, did not promote U4 snRNA release. This finding would imply Snu114 to act in a similar way to large GTPases involved in translation which change their conformation upon GTP hydrolysis leading to a downstream effect.

In a second study, contradictory results showed that binding of the nucleotide but not hydrolysis affects spliceosome activation (Small et al. 2006). In this study, tri-snRNP particles were purified by affinity capture via Prp28, and the RNA was extracted after incubation with different nucleotides. Activation was measured as the proportion of U4/U6 snRNA duplex versus free U4 and U6 snRNA on native PAGE. Similar results were obtained from experiments with a non-hydrolysable form of XTP/GTP. This finding proposes that Snu114 acts as a switch-like small GTPase, likely to be active in the GTP-bound form (Cherfils and Zeghouf 2013).

A third hypothesis derives from the structural information on tri-snRNP. The GTP nucleotide could have solely a stabilising effect on the fold of Snu114 and might not be subjected to hydrolysis or exchange during splicing.

2.2.1 Recombinant Snu114 binds GTP and has no intrinsic GTPase activity

In order to determine the nucleotide state of isolated Snu114, the bound nucleotide can be analysed from recombinantly expressed Snu114:Prp8-N by HPLC (high-performance liquid chromatography; for expression compare **subchapter 2.1**). The nucleotide bound in this particle might reflect the initial situation before activation, when Snu114 is bound in U5 snRNP but not assembled into a spliceosome yet.

The nucleotides, which were released by heat denaturation from the protein dimer, clearly co-elute with a GTP but not GDP nucleotide standard (**Figure 2-6**, GTP navy, released nucleotide grey). Native mass spectrometry confirmed the finding, identifying all Snu114:Prp8-N particles are bound to a molecule with the mass of a GTP nucleotide (in collaboration with J. Gault). Since no GDP-bound form nor the apo-form of Snu114 was found, which should be present if Snu114 would be able to hydrolyse GTP by itself, these results suggest that Snu114 lacks an intrinsic GTPase activity. Of course, it remains possible that an additional factor or a different spliceosome conformation can act as a GAP to facilitate hydrolysis in the natural context.

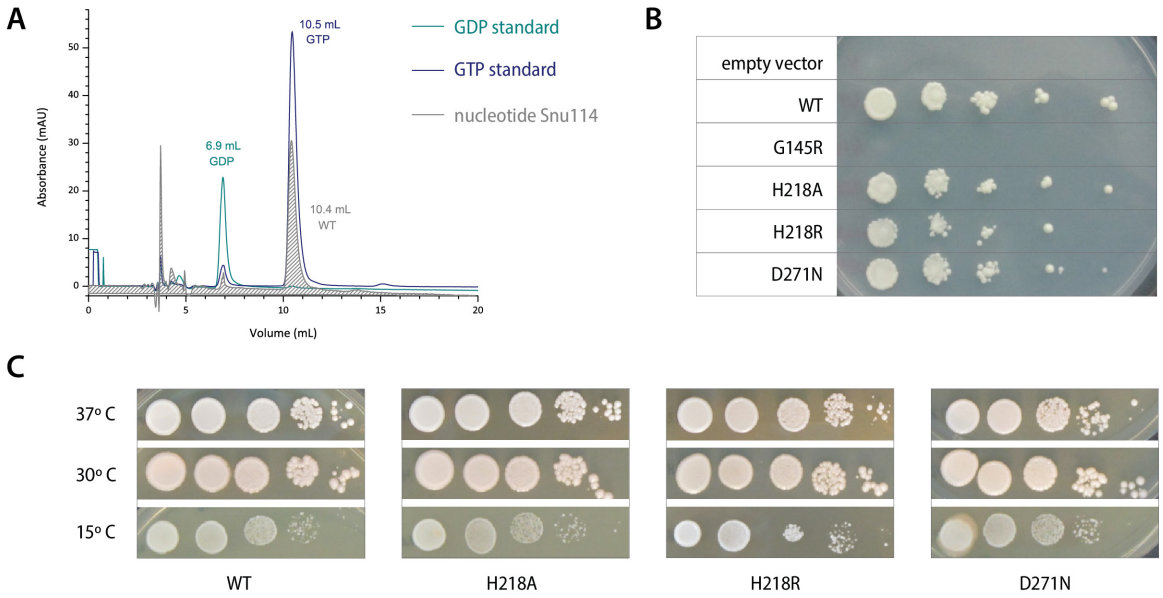


Figure 2-6: Snu114:Prp8-N binds GTP and nucleotide binding is required for yeast viability. (**A**) Released nucleotides from the dimer (Snu114:Prp8-N) coelute with the retention volume of GTP but not GDP standards in HPLC analysis. (**B**) Yeast viability assays. The Snu114-gene is replaced by a plasmid coding for Snu114 in the indicated form, grown at 30 °C on 5-FOA containing plates (**7.1.5 Materials and methods**). (**C**) Colonies were further grown on YEPD plates at the indicated temperature. (**B**) and (**C**) depict one out of three replicates.

2.2.2 Mutations of Snu114's catalytic residue have no effect on viability

For this subsection, I repeated yeast viability assays with Snu114 mutants, harbouring an alteration in their GTP binding pocket, to draw directly comparisons under the same experimental conditions. The translation factor EF-2, homologous to Snu114, coordinates the water molecule for GTP hydrolysis with a histidine residue and Snu114 also possesses a histidine at this position (**Figure 2-5**, His218). To analyse if the histidine has catalytic activity, meaning it can activate the water molecule necessary for hydrolysis, mutants with alterations at this position were compared in yeast viability assays. Altering His218 to an alanine or arginine, two amino acids that are not able to activate a water molecule, resulted in hardly any growth defect, and the mutants are barely temperature sensitive (**Figure 2-6 B** and **C**, this was also shown in (Nguyen et al. 2016)). In contrast, G145 which is located in the GTPase motif 1 and necessary to bind the nucleotide, is a lethal mutation when replaced by a long side chain such as arginine (**Figure 2-6 B**, and shown in (Fabrizio et al. 1997)). Taken together, this suggests that nucleotide binding by Snu114 is indeed essential for viability including splicing, but hydrolysis might not be.

In addition, Snu114-D271N (xanthine nucleotide-binding) was tested for viability. In my hands, the mutant did not show any growth defect at 30 °C but was mildly delayed at 14 °C (**Figure 2-6 C** last panel; compare (Bartels et al. 2003) in which a slow growth defect is noticed). Assuming that Snu114 requires a bound GTP for proper function in splicing, the normal phenotype of Snu114-D271N either

suggests that the level of endogenous XTP is high enough to saturate the minimally required amount of functional Snu114 molecules, or that the mutation itself is not completely specific and other nucleotides can replace the xanthine nucleotide in the binding pocket.

2.2.3 Spliceosomal complexes solved by cryoEM show Snu114 bound to GTP

Recently, structures of spliceosomal complexes which were captured at various states of splicing have been determined by cryoEM. Following Snu114, it joins the spliceosome bound in tri-snRNP to form complex pre-B (Bai et al. 2018), which transforms to the precatalytic complex B (Plaschka, Lin, and Nagai 2017) to then remodel to catalytically active complex B^{act} (Yan et al. 2016a) to accomplish the catalytic steps of splicing. In all these structures, the bound nucleotide of Snu114 is identified as GTP, casting another doubt for Snu114-mediated GTP hydrolysis in activation.
Nevertheless, it remains elusive what happens between these structural snapshots.

2.3 Spliceosome activation might happen independently of Snu114

Taken together, all these findings support the model of a constantly bound GTP molecule by Snu114, precluding a role for Snu114-catalyzed GTP hydrolysis in activation.
To prove the hypothesis and exclude the requirement of GTP hydrolysis during splicing, several experiments were planned, however, the project was discontinued in favour of *in vitro* spliceosome iCLIP (**chapters 3, 4 and 5**) before clear results were obtained. Some of the experimental approaches and their theoretical implications are discussed below:
We planned to take advantage of the orthogonal system, modifying the nucleotide binding site of Snu114 to perform experiments with xanthine nucleotides (XNP). To ensure specific and exclusive XNP-binding of the mutant, the nucleotide state of Snu114-D271N:Prp8-N should be analysed and binding to XNP tested. I was able to express the protein dimer with a 2.5-fold lower yield than the WT protein, indicating a less stable protein complex. Unfortunately, HPLC analysis of released nucleotides was not obviously interpretable. Binding of exogenous XTP to the complex could not be achieved either.
With the system in hand, nucleotide effects on unwinding were to be tested on purified spliceosomal complex B. In a first step towards this set up, Snu114-D271N-Strep was expressed from a low copy number plasmid in a strain expressing endogenous Brr2-proteinA for split tandem affinity purification. Incorporation for Snu114-D271N into Brr2 containing particles, however, is decreased resulting in a low yield. A yeast strain expressing Snu114-D271N endogenously might have offered an improvement.
To go further and examine if GTP is hydrolysed during splicing, a pseudo-haploid yeast strain expressing both Snu114-WT and Snu114-D271N-tag was generated. The WT copy allows the cells to

grow normally. Snu114-D271N containing extract was reported to have a temperature-dependent splicing defect (Bartels et al. 2003). This system would allow to test the effect of different nucleotides and to compare splicing efficiencies of the Snu114 WT and mutant proteins. After an *in vitro* splicing reaction, spliceosomes bound by the two different Snu114 copies can be separated by affinity capture to analyse their splicing phenotype on the basis of resulting RNA species. In addition, one should perform the experiment with suboptimal splicing substrates to investigate if Snu114 is required to maintain splicing fidelity.

In addition, incorporation of exogenous GTP/XTP into spliceosomes could be tested to reveal if nucleotide exchange happens during splicing (GDP release). Excluding GTP hydrolysis/exchange in all the listed experiments, would strongly lead to the conclusion that Snu114 is a pseudo GTPase. The bound GTP, which might be essential to maintain Snu114's three-dimensional structure, might remain from a time when the common ancestor of Snu114 and its homologue EF-2 was an active GTPase. Snu114 might have become part of the spliceosome due to its three-dimensional structure and lost its redundant GTPase activity.

At the end of this chapter, the question remains what the activating factor for Brr2 is, if Snu114 does not signal the helicase to unwind the U4/U6 snRNA duplex.

3 Establishing iCLIP for helicases in stalled and purified spliceosomes

As described in **1.5 Questions to be addressed in this project**, crosslinking and sequencing approaches are powerful techniques to identify the direct RNA targets of spliceosomal helicases. Even transient interactions with an RNA can be captured through the covalent tethering by UV crosslinking. To obtain a mechanistic understanding of the spliceosomal helicases, a defined system is needed. Therefore, I developed an iCLIP (individual nucleotide resolution UV cross-linking and immunoprecipitation) technique suitable for purified and stalled spliceosomal complexes. This chapter leads though the modifications and optimisations for *in vitro* spliceosomal iCLIP.

iCLIP was established to determine the interaction of an RBP and its RNA targets at nucleotide resolution in a genome-wide manner (Konig et al. 2010; Huppertz et al. 2014). The pipeline established by the J. Ule laboratory is shown in **Figure 3-1**. Living cells or extracted tissue are irradiated with 254 nm ultraviolet (UV) light to stimulate the intrinsic crosslinking activity of RNA resulting in covalently connected RNA-RBP complexes. After cell lysis, a controlled RNA digestion is performed to obtain crosslinked RNA fragments of lengths compatible with next-generation sequencing (NGS). The RNA-RBP complex is captured by immunoprecipitation using a protein-specific antibody coupled to beads while stringent washes remove non-covalent interactions. Depending on the RNase used, the 3'-end of the RNA may need to be dephosphorylated to allow ligation to a universal DNA adapter (IR-L3 adapter, compare **Figure 3-1**), which provides the annealing sequence for later reverse transcription. In addition, the adapter is conjugated to an infrared (IR) fluorophore facilitating visualisation of the RNA-RBP complex when blotted onto nitrocellulose membrane after separation by SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis) (Zarnegar et al. 2016). With the correct RNase concentration, a smear starting strongly around the size of the protein fading towards higher molecular weight is produced. Too high RNase concentrations will lead to a distinct band around the size of the protein. On the other side, concentrations lower than the optimum will result in long RNA species which cannot enter the gel and result in a diffuse signal. When multiple RBPs bind the same RNA as the target protein, proper RNase digestion and size-selection by SDS-PAGE will help to separate these. The RNA-RBP complexes are subsequently extracted by pro-

teinase K digestion leaving a peptide at the crosslinked site. Library preparation for sequencing begins with reverse transcription (RT) using primers, which anneal to the IR-L3 adapter and contain the P3 Solexa sequence. In reverse order, a hexa-ethyleneglycol spacer separates the P5 Solexa sequence, and a unique identifier (barcode) at the 5'-end of the phosphorylated primer. P3 and P5 Solexa sequences are the flow cell attachment site in Illumina NGS. In addition, each RT primer introduces a randomised sequence component to identify PCR (polymerase chain reaction) duplicates during later analysis. The RT reaction will truncate just prior to the crosslinked site. The resulting cDNAs are circularised such that the truncation site is directly linked to the RT-primer barcode sequence. After amplification with P3 and P5 Solexa primers, the linear cDNA is gel purified and quantified before loading onto the Illumina platform for NGS.

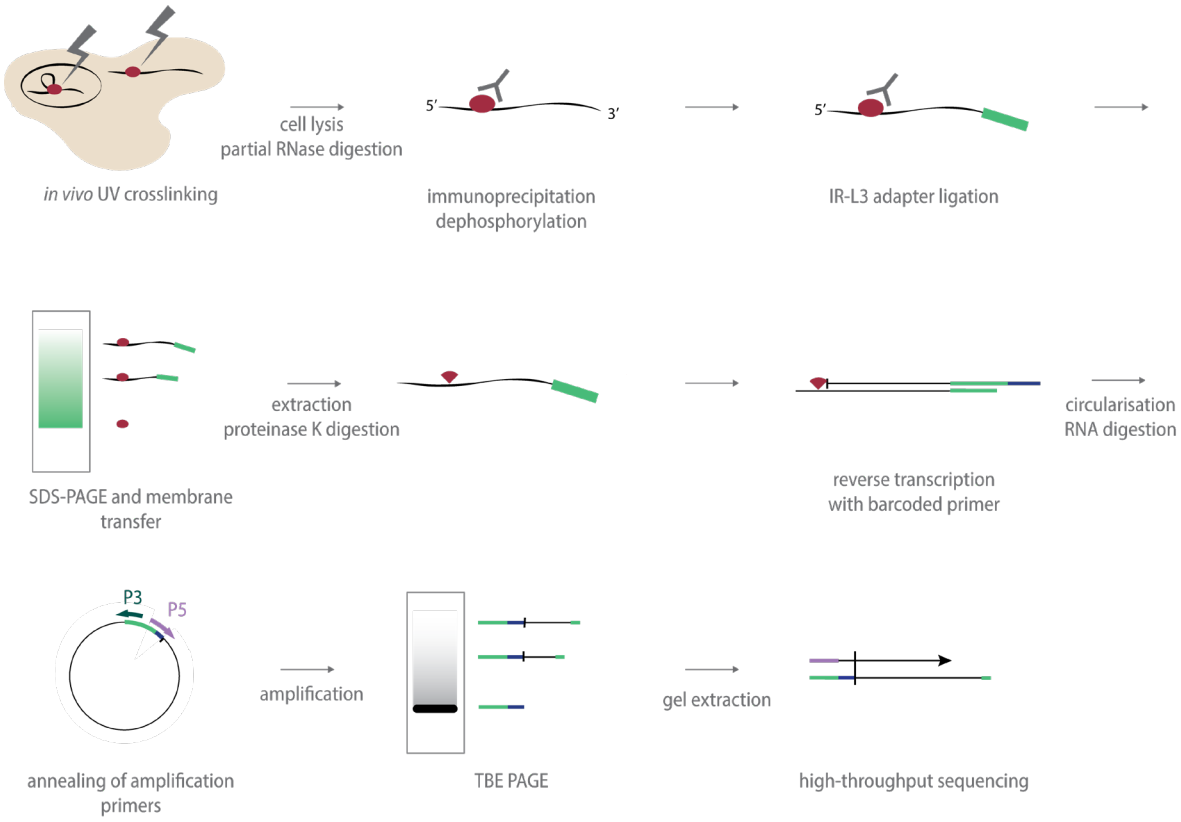


Figure 3-1: Workflow of the conventional iCLIP protocol (revised by Ule lab from (Huppertz et al. 2014)). Cells are irradiated with 254 nm UV light to obtain covalent crosslinks between RBPs and their target RNAs. After cell lysis, fragmentation of the RNA generates suitable lengths for Illumina sequencing. While capturing the protein of interest by immunoprecipitation, the RNA 3'-end is dephosphorylated to allow IR-L3 adapter ligation. The covalently linked RNA-RBP complexes are separated by SDS-PAGE, blotted onto nitrocellulose membrane, and extracted by proteinase K digestion. Reverse transcription truncates at the crosslink site occupied by the remaining protein peptide. Barcoded primers will identify reads from different experiments and the random sequence allows for later identification of reads derived from PCR duplicates. The cDNA is circularised, and the RNA template digested. After further purification of the cDNA, it is amplified directly with P3 and P5 Solexa primers. The resulting library is purified on a 6 % TBE (Tris-borate-ethylenediaminetetraacetic acid) PAGE to separate cDNA from free IR-L3 adapter and primer artefacts. After gel extraction and quantification, the libraries from different experiments are multiplexed and loaded onto the Illumina HiSeq 2500 system for single-read sequencing.

The main advantage of iCLIP is that the precise truncation event occurs at the crosslink site in about 80-100 % of reverse transcription reactions; this allows for identification of the interaction site with nucleotide resolution (Haberman et al. 2017). Other CLIP-based methods such as HITS-CLIP/CLIP-Seq (high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation) and PAR-CLIP (photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation), in which crosslinking events are identified from read-through events, lose information and accuracy during the library preparation step due to a more laborious adapter configuration. iCLIP is constantly being modified and developed further, for example by isolating the RNA-RBP complex without its visualisation by SDS-PAGE in eCLIP (enhanced CLIP), therefore saving time and experimental steps (Van Nostrand et al. 2016).

By providing the precise location of where an RNA-binding protein binds on its target RNA, iCLIP was considered to be an ideal starting point for the development of *in vitro* spliceosomal helicase iCLIP. In addition to adjusting RNase amounts to obtain suitable lengths for next-generation sequencing, most other steps had to be optimised and adapted to an *in vitro* system. In contrast to irradiated human cells, were the UV dose is held constant throughout the protocols, purified spliceosomes derived from *in vitro* splicing reactions in yeast are the starting material here. This set-up allows for variation of the pre-mRNA substrate, the addition of dominant negative mutant helicases, or the depletion of specific spliceosomal factors. Basically, incubation of pre-mRNA substrate in yeast extract under splicing conditions will lead to assembled spliceosomes (**Figure 3-2 A**). The assembled complexes can then be purified using size selection methods, or affinity strategies on either the RNA substrate (MS2-stem loops shown in **Figure 3-2**) or spliceosomal proteins. As described in the introductory section **1.4** using modified pre-mRNA substrates or dominant negative proteins, which stall the splicing process at a specific step in combination with capture via step-specific factors, produces homogeneous spliceosome species. As a flexible alternative to using protein specific antibodies for immunoprecipitation during the iCLIP procedure, we chose to add epitope tags to the target protein via homologous recombination.
Initially, Prp16 helicase in spliceosomal complex C served as the first target protein to develop the method. In addition, we decided that the RNA contacts of SmB protein should be determined in complex C, serving as a positive control. In a second step, additional spliceosomal helicases will be targeted by transferring and adjusting the method (compare **chapter 5**).

There are several possible ways to stall an *in vitro* splicing reaction at the complex C stage (introductory section **1.4**), for example by the addition of dominant negative Prp16 protein (Prp16-DN) (Schneider, Hotz, and Schwer 2002) or a pre-mRNA substrate with a 3'SS mutation that causes a

second step defect (Vijayraghavan et al. 1986) (**Figure 3-2 B** and **C**). In the following chapters **3** and **4**, a complete permutation of these two strategies obtains a full dataset for comparison: WT protein and mutant substrate, mutant protein and mutant substrate, mutant protein and WT substrate.

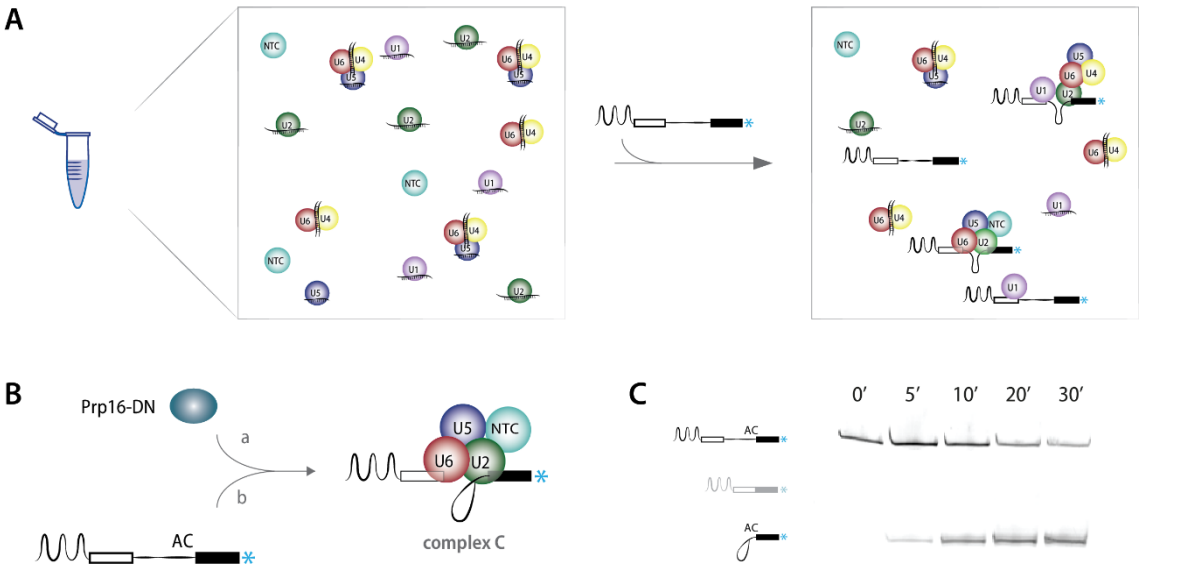


Figure 3-2: Schematic description of *in vitro* splicing. (**A**) Splicing extract contains pre-assembled snRNPs and other essential spliceosomal proteins. Incubation with a pre-mRNA substrate (carrying three MS2-stem loops at the 5'-end and a fluorophore at the 3'-end which is indicated with a *) starts assembly of spliceosomes. (**B**) An *in vitro* splicing reaction can be stalled at the complex C stage by either the addition of (a) dominant negative Prp16 (Prp16-DN) or (b) pre-mRNA with a mutation at the 3'SS whereby AG is replaced by AC. (**C**) 10 % denaturing PAGE shows a time course of a splicing reaction for *UBC4*_AC accumulating intron-lariat intermediate. Spliced RNA would migrate between pre-mRNA and intron-lariat intermediate.

Establishing an iCLIP method for Prp16 in complex C will allow us to better understand the Prp16 RNA remodelling mechanism and will serve as a starting point for adapting the approach to the analysis of other spliceosomal helicases. Going one step further, it is essential to collect data points for the position of Prp16 at the stalled complex C step before the system can be altered to conduct time-resolved studies which will follow the helicase binding path on its substrate over time.

3.1 Purified spliceosomal complex C is the starting material for *in vitro* spliceosome iCLIP

To obtain the stalled and homogenous starting material for the iCLIP method, spliceosomal complex C was purified from a 1.5 mL *in vitro* splicing reaction. The *UBC4*-derived pre-mRNA substrate was adapted from cryo-EM studies (Galej et al. 2016) to facilitate comparisons to the three-dimensional structure. The sequence of the last 25 nt of exon 1, the 95 nt intron and the first 25 nt of the second exon were extended by three MS2 stem loops on the 5'-end (referred to as 3xMS2_*UBC4*_short). The mutated version, in which AG|AG is altered to AC|AC (where | indicates

the 3'SS), stalls spliceosomes at the complex C step (referred to as 3xMS2_*UBC4*_AC_short) (**Figure 3-3 A**). Since another AG directly follows the 3'SS in the *UBC4* gene, the AC|**AC** mutation was necessary to avoid redirecting splicing to this second site. Prp16-G378A was previously described as a mutant with an impaired ATPase activity and strong dominant negative effect (Schneider, Hotz, and Schwer 2002). Since it is unclear whether spliceosomes stalled by the substrate arrest in the same way as by the ATPase-deficient helicases, Prp16-G378A was included in the study. A calmodulin-binding peptide (CBP) and octa histidine-tagged version (rPrp16-G378A) was purified recombinantly from yeast, which replaces the wildtype protein when added in excess to the splicing extract before the reaction.

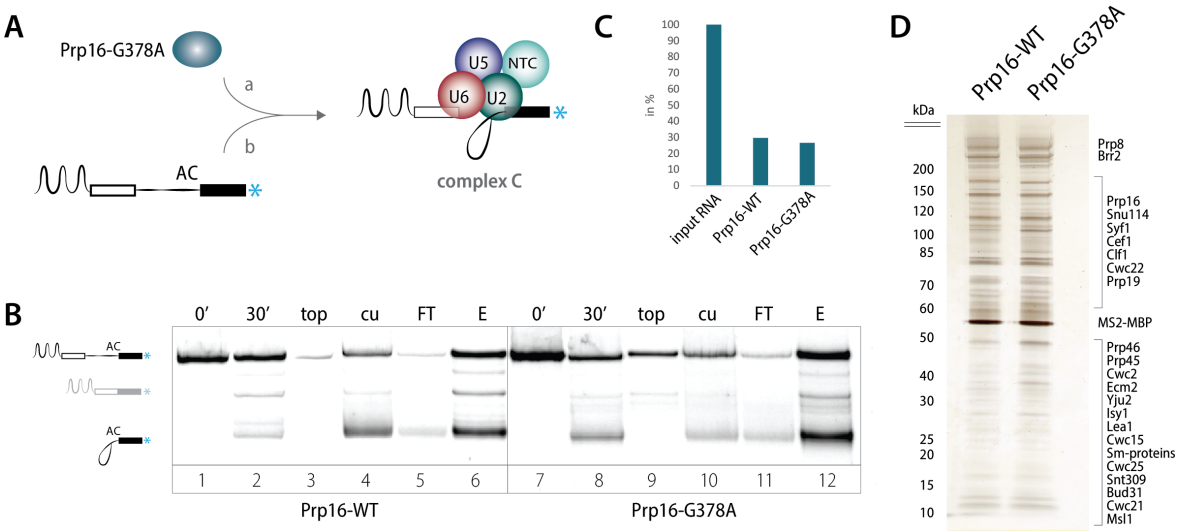


Figure 3-3: Spliceosome purification. (**A**) Schematic presentation of the strategies for spliceosomal complex C purification. The splicing reaction can either be stalled by adding dominant negative Prp16 (in this study: Prp16-G378A) or by using a pre-mRNA substrate with the 3'SS mutation AC (in this study: 3xMS2_*UBC4*_AC). The latter strategy can be combined with either Prp16-WT or Prp16-DN. (**B**) RNA species are visualised by the 3'-Cy5-label on a 10 % denaturing PAGE: before splicing (lanes 1 and 7) after 30 min incubation (lanes 2 and 8), top (lanes 3 and 9) and cushion (lanes 4 and 10) after centrifugation through the glycerol cushion, flow-through (lanes 5 and 11) and elution (lanes 6 and 12) of the amylose capture. The RNA band migrating between pre-mRNA and intron-lariat intermediate derives most likely from a splicing event using a cryptic site. Left: extract containing Prp16-WT, right: extract supplemented with dominant negative rPrp16-G378A. (**C**) Representative amounts of purified spliceosomes compared to input pre-mRNA measured by fluorescence of the Cy5-labelled RNA substrate. (**D**) Silver stained PAGE showing protein composition in the amylose eluate. Prp8, Brr2, and MS2-MBP can be identified from their migration pattern. The region, in which several other complex C proteins migrate, is roughly indicated but cannot be assigned with certainty due to a mixture of spliceosomes before and at the first catalytic step.

After spliceosome assembly, a glycerol cushion separated free pre-mRNA and small particles in the top from the spliceosomes, which accumulate in the higher-viscosity cushion (**Figure 3-3 B**, compare lane 3 to 4, and lane 9 to 10). After amylose capture of the assembled spliceosome via the MS2-MBP (maltose-binding protein) system, about a third of the starting pre-mRNA substrate could be eluted as a mixture of assembled spliceosomes before and at the first catalytic stage bound to

pre-mRNA and intron-lariat intermediate, respectively (**Figure 3-3 B** lanes 6, 12, **C** and **D**). Introducing an ATP-depletion step by adding glucose after the splicing reaction increased the purification efficiency of the spliceosome (representative gel in **Figure 3-3 B**). Addition of glucose results in the depletion of ATP in the extract as the glucose is phosphorylated by hexokinases utilizing ATP. Removing the ATP from the extract might have increased the accumulation of stalled complexes that otherwise would have been subject to disassembly. Although leading to a mixture of spliceosomes as indicated by the presence of pre-mRNA and intron-lariat-intermediate (**Figure 3-3 B** lanes 6, 12 and **D**), this two-step purification was considered as sufficient at this stage. The subsequent capture on Prp16, a step 1 specific splicing factor, during the iCLIP procedure, ensures a homogenous population representing the spliceosomal complex C stage (**chapter 3.2**).

The data in this chapter is derived from spliceosomes purified from extract containing either endogenous Prp16-WT or supplementary rPrp16-G378G assembled on 3xMS2_*UBC4*_AC_short. In **chapter 4** spliceosomes containing rPrp16-G378A assembled on 3xMS2_*UBC4*_WT were purified to use in *in vitro* spliceosome iCLIP.

3.2 Helicases are captured and can be monitored by an affinity tag

3.2.1 Polyhistidine-tags have advantageous properties but produce high background signal

In the conventional iCLIP protocol, the target protein is captured by a protein-specific antibody. The spliceosomes containing the target protein for *in vitro* spliceosome iCLIP, however, originate from yeast which can be genetically modified to introduce C-terminal affinity tags by simple homologous recombination (**7.1.4 Materials and methods**). Affinity tags exhibit various properties such as sustaining binding capacity under stringent conditions like the commonly used polyhistidine-tag. These properties seemed beneficial to capture solely the crosslinked RNA-RBP complex under denaturing conditions and reducing unspecific background. The Prp16-His_8 allele supported normal growth of the yeast strain. As the only copy of the essential *PRP16* gene, this directly demonstrates intact function of the modified protein. rPrp16-G378A was similarly expressed with a C-terminal octa histidine-tag.

Ideally, the target protein can be monitored via its affinity tag during the multi-step procedure of *in vitro* spliceosome iCLIP, however, a high background signal was obtained when probing for the polyhistidine-tag by western blot in whole cell extract and even after capture with TALON™ beads. Only the purified rPrp16-G378A-His showed a distinct signal (data not shown). For this yeast-derived system, the polyhistidine-tag was unsuitable to ensure protein detection throughout the iCLIP preparation.

3.2.2 3xFLAG-tag enables helicase monitoring throughout the iCLIP procedure

As described before, visualisation of the protein during the immunoprecipitation would facilitate optimisation and troubleshooting of the iCLIP procedure. In a second approach, antiserum against Prp16 was explored for western blot detection. Although this serum was successfully used in immunoprecipitations by (Schwer and Guthrie 1992a), the background was comparably high as probing with anti-His antibody in western blot. Detection and visualisation of the genetically introduced C-terminal single FLAG-tag by anti-FLAG antibody in western blot was also ineffective (data not shown). Finally, a 3xFLAG-tag, introduced by homologous recombination at the C-terminus of Prp16, produced a clean and distinct signal in western blot for both endogenous Prp16-3xFLAG-His and overexpressed rPrp16-G378A-3xFLAG-CBP-His in yeast whole cell protein extract (**Figure 3-4 A**).

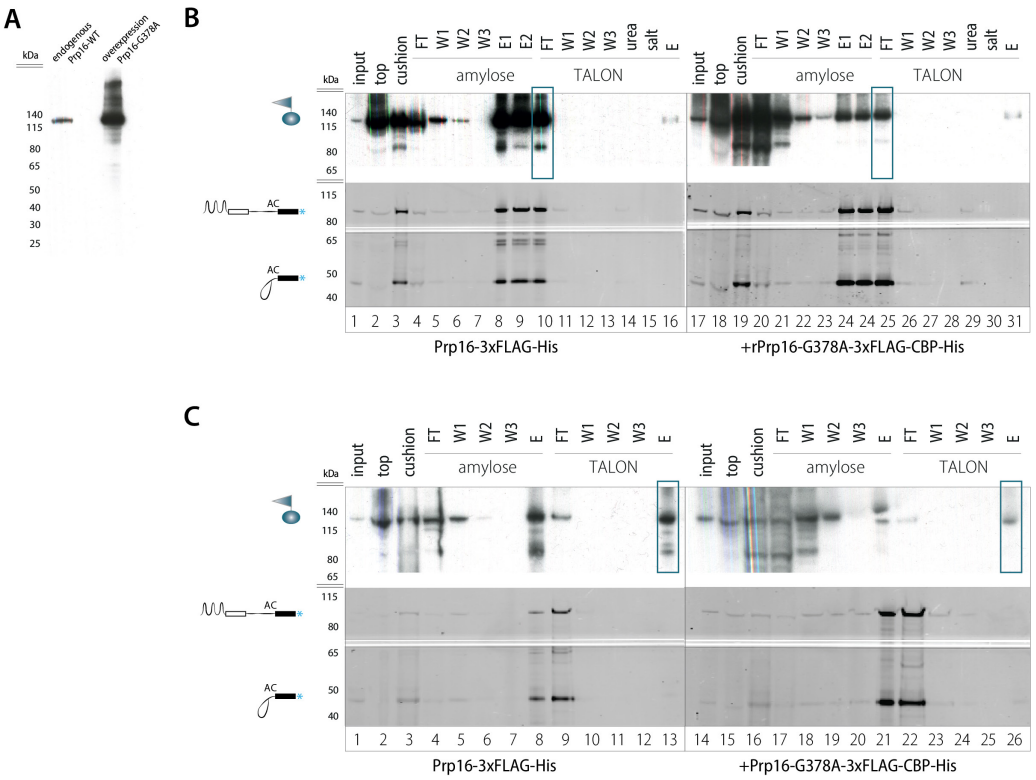


Figure 3-4: Troubleshooting benefits from 3xFLAG-tag. (**A**) Western blot probing with anti-FLAG antibody for Prp16-3xFLAG-His in whole protein extract (left; a 10-fold excess of total protein amount was loaded) and for Prp16-G378A-CBP-3xFLAG-His in extract made from yeast cells overexpressing the protein (right: cells were induced with galactose and harvested at a similar OD compared to the cells used for the left lane; the used plasmid is listed in **7.1.6 Materials and methods**). (**B**) and (**C**) Spliceosomal complex C purification and TALON™ capture of polyhistidine-tagged target protein showing inefficient versus efficient TALON™ capture under (**B**) native and (**C**) denaturing conditions containing 6 M urea. Left: derived from splicing extract with endogenously tagged Prp16-3xFLAG; Right: derived from WT splicing extract supplemented with rPrp16-G378A-3xFLAG-CBP-His. Top: Western blot probing for 3xFLAG-tag. The expected retention position for the FLAG-tagged protein is indicated with a cartoon showing a cyan sphere with a flag; Bottom: SDS-PAGE (original gel) visualising Cy5-labelled RNA species. The expected retention positions for pre-mRNA and the intron-lariat intermediate are indicated with cartoons. Lanes: input (5-fold dilution), top (top) and bottom part (cu) of the glycerol cushion, FT (flow-through), W1-3 (washes), E1-2 (elution from amylose), urea (6 M urea wash), salt (1 M NaCl wash), E (elution from TALON). In both (**B**) and (**C**) left and right are derived from separate gels.

3.2.3 Efficient capture of C-terminally tagged Prp16 requires prior denaturation

By carrying the 3xFLAG-tag, Prp16 becomes trackable throughout the whole procedure (compare **3.2.2**). Thus, spliceosome purification followed by the immunoprecipitation steps of the protein were monitored for troubleshooting. Most Prp16 helicase does not assemble into spliceosomes, consistent with a much higher proportion of anti-FLAG signal compared to the RNA amount in the top part of the cushion (lane 2) than in the cushion (**Figure 3-4 B**, lane 3). While almost all RNA is retained during amylose capture, a comparable higher amount of Prp16 is lost while binding to (lane 4) and washing (lanes 5-7) the beads. This confirms Prp16 to be a weakly or transiently bound component of the spliceosome. However, in the purified spliceosome, Prp16 is present confirming successful purification of spliceosomes containing complex C. Contrary to expectation, the protein is hardly captured by TALON™ subsequently, indicating the C-terminal tag of Prp16 is inaccessible when assembled into spliceosomes (**Figure 3-4 B**, lanes 10 to 16).

The protein capture was clearly improved when the spliceosomes were denatured with 6 M urea or 5 M guanidinium hydrochloride prior to loading onto TALON™ beads. Comparing the amounts of unbound (lane 9) and captured Prp16 (lane 13), the pulldown efficiency was improved significantly under urea conditions (**Figure 3-4 C**). Denaturation with guanidinium hydrochloride enhanced binding too, however less effectively (data not shown).

Comparing these results for WT splicing extract supplemented with rPrp16-G378A-3xFLAG-CBP-His and experiments performed on endogenously tagged Prp16, the conclusions are the same (left and right panels, **Figure 3-4 B** and **C**). A higher amount of rPrp16-G378A is present at the start of the procedure, and most excess is lost during the early purification steps. The amount of rPrp16-G378A, which is added to the extract before the splicing reaction, was therefore adjusted to reach the same level as the WT protein during the immunoprecipitation step.

3.3 Optimisation of UV-crosslinking and enzymatic reactions for *in vitro* spliceosome iCLIP

3.3.1 UV irradiation can covalently link Prp16 and RNA in purified spliceosomes

Despite successful TALON™ capture of Prp16-His under denaturing conditions, after a whole iCLIP procedure, the expected RNA-Prp16 crosslinked species were still not appearing on SDS-PAGE (data not shown, expected result shown in **Figure 3-1**). In the unsuccessful test *in vitro* iCLIP experiment, UV and RNase amounts were adapted from the standard iCLIP procedure irradiating with 1,000 x 100 µJ/cm^2 UV light and testing two different concentrations of RNase I.

RNA-protein crosslinking of the target protein is essential to pursue this method, however some RBPs fail to covalently link to their target RNA, as nucleobases and the target protein need to be in

an optimal conformation to allow photo-activated formation of a covalent bond. Insufficient crosslinking of Prp16 to its target RNA was considered as an explanation and therefore purified complex C was targeted for UV-titration. The protein migration shifts upwards while the intensity of the band corresponding to free Prp16 decreases with increasing UV exposure (**Figure 3-5**), demonstrating successful crosslinking of Prp16 and RNA. The labelled pre-mRNA and intron-lariat-intermediate likewise fade out with increasing UV irradiation, indicating the RNA substrate crosslinks to proximal spliceosomal proteins, producing new RNA-protein species of reduced gel mobility. The crosslinking efficiency for Prp16-WT and rPrp16-G378A seems comparable (**Figure 3-5** compare **A** and **B**).

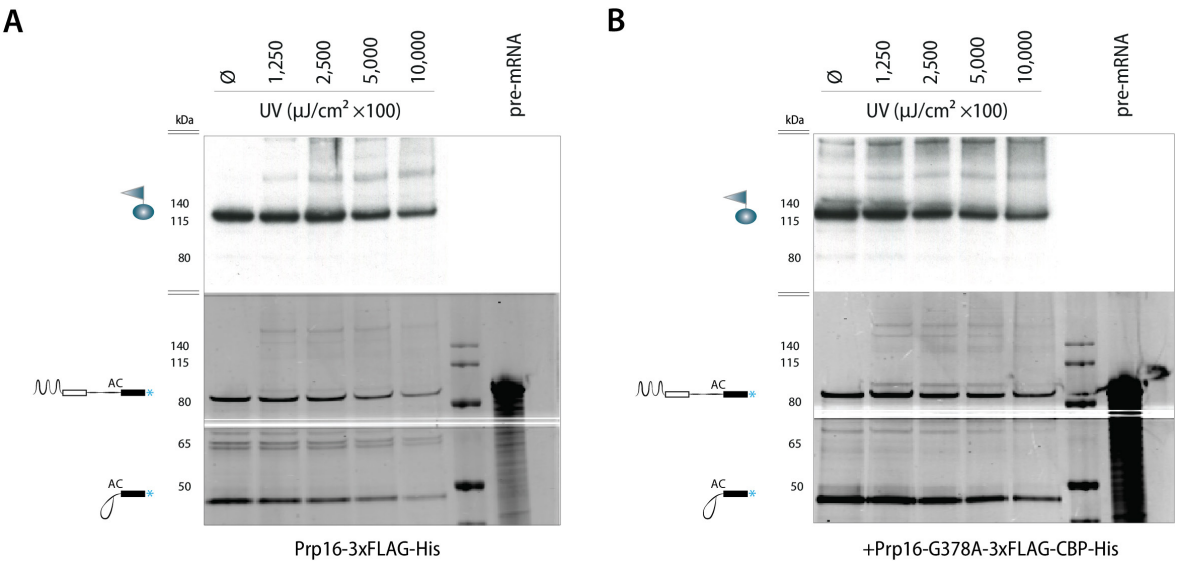


Figure 3-5: Prp16 is crosslinked to RNA by UV irradiation. Irradiation of purified complex C with different amounts of UV light including a negative control (Ø), 1,250, 2,500, 5,000, and 10,000 x 100 µJ/cm^2. Top: Western blot probing with anti-FLAG antibody; Bottom: SDS-PAGE (original gel) scanned for Cy5-labelled RNA species. (**A**) Spliceosomes purified from extract expressing endogenous Prp16-3xFLAG-His and (**B**) WT extract supplemented with rPrp16-G378A-3xFLAG-CBP-His.

Ideally, one aims to achieve a maximum of one crosslink event per molecule pair. Thus, all crosslinked peptides will generate truncation events in later reverse transcription. The UV amount used here, however, seems to crosslink the majority of protein-RNA complexes indicating 'over crosslinking'. Therefore, UV irradiation had to be precisely adjusted, which is discussed in **3.3.3.**

3.3.2 Optimisations for RNase digestion, dephosphorylation and IR-L3 adapter ligation

The three enzymatic reactions that are necessary for efficient IR-L3 adapter ligation were tested in a defined setup to carefully troubleshoot each step, which could prevent formation of the RBP-RNA-adapter complex.

In *in vitro* spliceosome iCLIP, the RNA is cleaved during the immunoprecipitation while the protein is bound to beads. In conventional iCLIP, however, RNase digestion is performed in cell lysate,

where 2.5 and 0.5 units RNase I per 1 mL of lysate are used for high and low RNase conditions, respectively. RNase I cleaves with similar efficiency after all four nucleotides and therefore introduces less bias compared to other RNases. To get a first impression of RNase I digestion on a more purified target such as the spliceosome, Cy-5 labelled pre-mRNA was digested by four different concentrations of RNase I in a 10-fold dilution series (compare **Figure 3-6 A**, 'RNase I digestion'), whereby a continuous range of different size fragments is obtained (**Figure 3-6 B**). In addition, the two lower concentrations give an indication for a suitable number of units for the *in vitro* spliceosome iCLIP experiment. Similarly, S1 nuclease was tested, however, the digestion neither led to a continuous range of lengths nor showed clear concentration-dependency (data not shown).

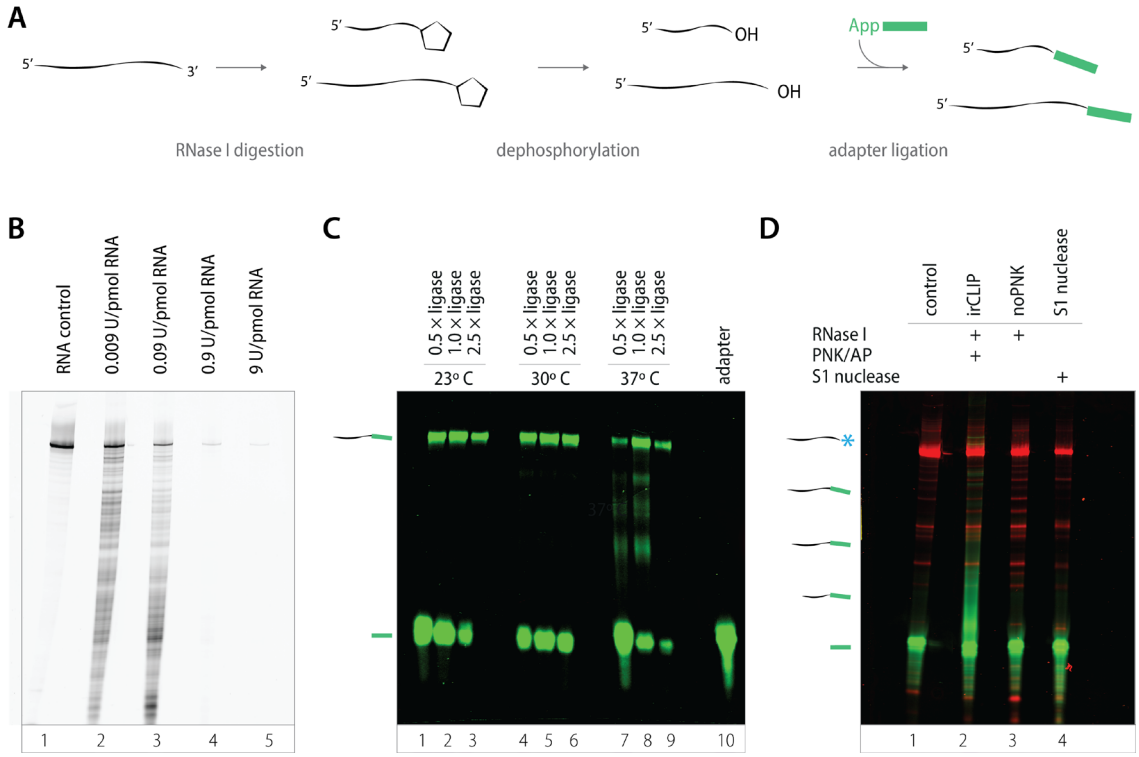


Figure 3-6: Optimisations for enzymatic RNA digestion and IR-L3 adapter ligation. (**A**) Schematic diagram of RNA fragmentation with RNase I and adapter ligation. RNase I leaves, in contrast to some other RNases, a 2',3'-cyclic phosphate at the 3'-end (depicted as a pentagon) which has to be dephosphorylated before the free OH-group can react with the pre-adenylated IR-L3 adapter (depicted as a green bar). (**B**) Cy5-labelled pre-mRNA substrate (*UBC4*_AC_short) was incubated with different amounts of RNase I in 200 µL RNase buffer as indicated (RNA amount: 1.7 pmol). The reaction was stopped and deproteinised before visualising on an 8 % denaturing PAGE. (**C**) *In vitro* transcribed U5 snRNA A was incubated with pre-adenylated IR-L3 adapter under ligation conditions with the indicated amount of RNA ligase I at the indicated temperature for 16 hours. The reaction was stopped and deproteinised before separating on an 8 % denaturing PAGE. Molecules attached to the IR-L3 adapter appear green in the scan. Free IR-L3 adapter was run in lane 10, and its migration position is indicated with a green bar. Ligated products are indicated with a green bar attached to a black line. (**D**) 7.5 pmol Cy5-labelled pre-mRNA substrate was incubated with different RNases as indicated in the figure. After the reaction was stopped, deproteinised, and precipitated, the RNA was incubated with (lane 2) or without (lanes 3 and 4) T4 PNK and fastAP. After another round of deproteinization and precipitation, the RNA was incubated with pre-adenylated IR-L3 adapter and RNA ligase I at 23 °C for 16 hours. Final deproteinization and precipitation was done before separation on an 8 % denaturing PAGE. Molecules with a Cy5-label (starting molecule and products of RNase digest) and molecules attached to the IR-L3 adapter appear red and green, respectively.

In a second step, IR-L3 adapter ligation was tested on free RNA. Attaching the pre-adenylated IR-L3 adapter to the free 3'-OH-group of the target RNA by RNA ligase I is supposedly one of the most inefficient steps in iCLIP. In this test, the RNA possesses a free 3'-OH-group, and therefore the ligation efficiency can be tested without prior modification of the 3'-end (compare **Figure 3-6 A**, 'adapter ligation'). An increase in temperature leads to a slight increase in IR-L3-adapter-ligated RNA (**Figure 3-6 C**). The nucleotide species longer than free linker and shorter than correctly linked RNA-DNA could either result from degradation at higher temperatures or the formation of IR-L3 adapter concatamers. In theory, the biotin label at the 3'-end of the IR-L3 adapter should prevent concatamer formation. A higher concentration of RNA ligase I clearly leads to more efficient ligation (**Figure 3-6 C**).

In the full iCLIP experiment, IR-L3 adapter ligation follows RNase digestion. As RNase I leaves a 2',3'-cyclic phosphate at the 3'-end, this linkage has to be opened and the phosphate removed before ligation can occur (**Figure 3-6 A**). The reaction is most successful, when a combination of polynucleotide kinase (PNK) and alkaline phosphatase (AP), as described for iCLIP techniques in (Van Nostrand et al. 2016), removes the cyclic phosphate before IR-L3 adapter ligation (**Figure 3-6 D**, compare lane 2 to 1 and 3). S1 nuclease, which has the advantage of leaving a free 3'-OH after cutting, making additional dephosphorylation unnecessary, could not be optimised to lead to similarly efficient IR-L3 adapter attachment (**Figure 3-6 C**, lane 4).

As a result, RNase I was chosen to be further optimised for the controlled RNA digestion on crosslinked protein-RNA complexes, while high concentration RNA ligase I was used at 23 °C during the *in vitro* spliceosome iCLIP procedure for IR-L3 adapter ligation.

3.3.3 Adjusted UV dose and RNase I amount lead to appropriate lengths of crosslinked RNA

It is hard to predict the required amount of target protein for *in vitro* spliceosome iCLIP from conventional iCLIP, where a third of a 10-cm-diameter plate with adherent cells is the starting material. Therefore, yields of Prp16 crosslinked to RNA derived from spliceosomes, which were purified from a 1.5 mL splicing reaction, were compared by the amount of fluorescence via the Cy5-labelled RNA. The amount of input pre-mRNA was assessed against three conditions after TALON™ capture, whereby the first was not crosslinked, the second irradiated with the high UV dose used before (compare **Figure 3-5**), and the last condition both irradiated and treated with 0.3 units RNase I. RNase digestion releases the Cy5-label indicated by the fluorescence, which drops to the baseline level of the non-crosslinked condition (**Figure 3-7 A**, compare noUV and UV, RNase). Overall, about 5 fmol of RNA crosslinked to Prp16 are estimated to be captured, which should lead to a sufficient amount of cDNA in the final sequencing library (Zarnegar et al. 2016).

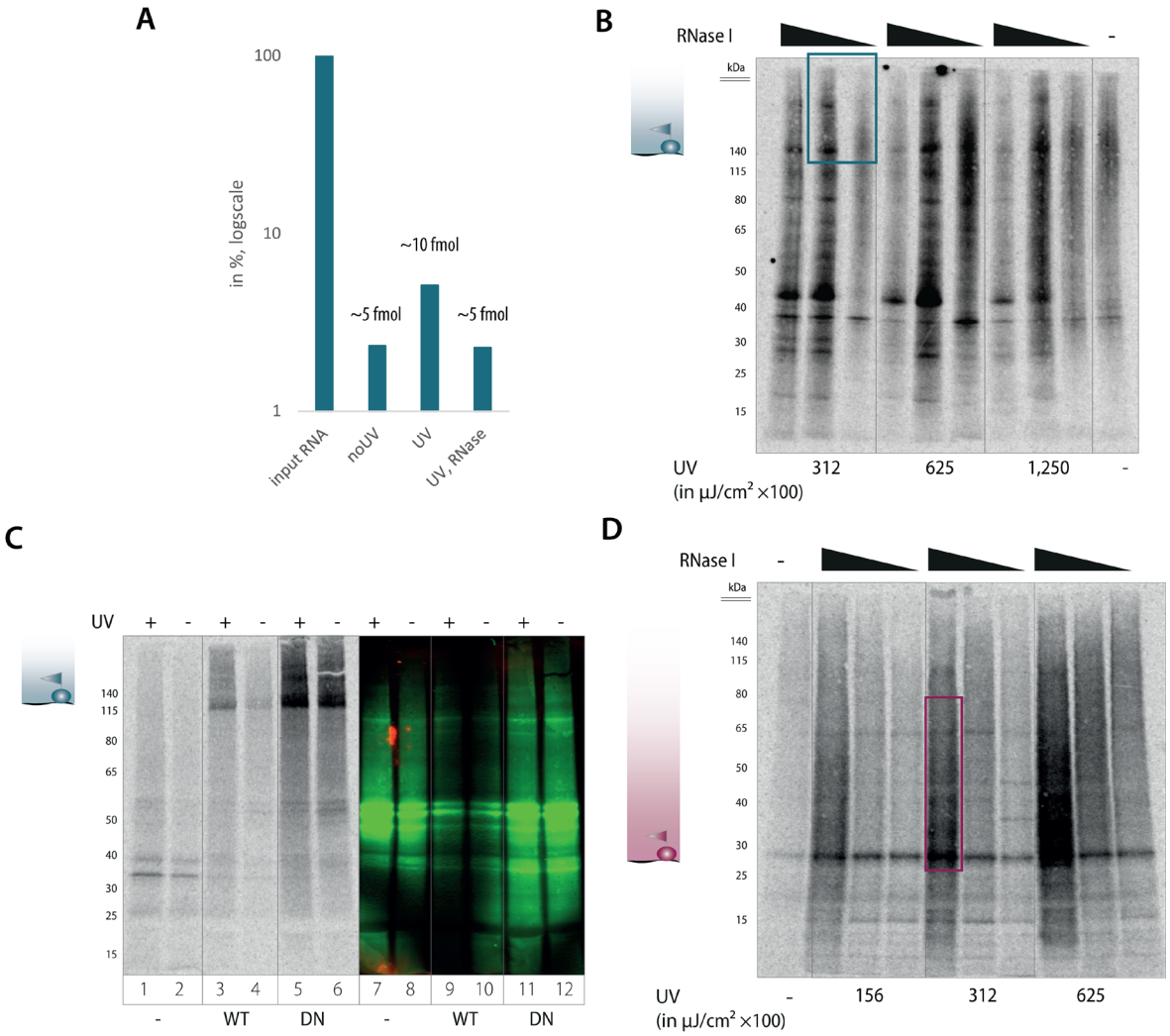


Figure 3-7: Optimal UV and RNase dose produces RNA fragments of suitable lengths. (**A**) Fluorescence quantification of input pre-mRNA substrate of a 1.5 mL splicing reaction compared to the amounts measured after TALON™ pulldown without UV irradiation (noUV), with prior UV irradiation (UV), and prior UV irradiation followed by RNase digestion (UV, RNase). The calculated RNA amounts are denoted above the bars. (**B**) Parallel fine-tuning of UV irradiation and RNase I digestion. Purified spliceosomal complex C was irradiated with 312, 625, and 1,250 x 100 µJ/cm^2 UV light, followed by denaturation, TALON™ capture, and digested with 1, 0.1, and 0.01 units RNase I. Subsequent γ-^{33}P-ATP radiolabelling allowed visualisation after SDS-PAGE separation and transfer onto nitrocellulose. The data derived from spliceosomes containing Prp16-WT; the highlighted lanes show a range for optimal conditions which was obtained likewise for rPrp16-G378A containing spliceosomes (not depicted). (**C**) Blot, showing radiolabelled RNA (left), IR-L3 adapter (right) after an *in vitro* spliceosome iCLIP experiment using FLAG capture including controls (lanes 2, 4, 6, 8, 10 and 12 for non-irradiated samples; lanes 1, 2, 7 and 8 for samples without a FLAG-tagged protein). Purified complex C containing either Prp16-WT (WT) or rPrp16-G378A (DN) were irradiated with 312 x 100 µJ/cm^2 UV light and treated with 0.5 units RNase I during FLAG capture. (**D**) Parallel fine-tuning of UV irradiation and RNase I digestion. Purified spliceosomal complex C was irradiated with 156, 312, and 625 x 100 µJ/cm^2 UV, followed by denaturation, FLAG capture, and digestion with 5, 0.5, and 0.05 units RNase I. Subsequent γ-^{33}P-ATP radiolabelling allowed visualisation after SDS-PAGE separation and transfer onto nitrocellulose. The data derived from spliceosomes containing endogenous SmB-3xFLAG-His, the highlighted lane shows the optimal condition. (**B**, **C**, **D**) The expected retention position for the FLAG-tagged protein is indicated with a cartoon showing a cyan sphere with a flag and a coloured gradient representing crosslinked RNA.

The amount of UV irradiation and RNase I were further varied in parallel and the RNA species obtained were visualised by radiolabelling. The inefficient steps of IR-L3 adapter ligation can thereby be avoided. Radiolabelled γ-^{33}P-ATP was used for enzymatic labelling of the 5'-end of any crosslinked RNA by T4 PNK during TALON™ capture of Prp16-His. This ATP-labelling reaction was optimised to be performed in 5 min. The visualised RNA suggests an optimal UV irradiation with 312 x 100 µJ/cm^2 combined with 0.1 to 0.01 units RNase I per experiment (**Figure 3-7 B**, shown for Prp16-WT). While increasing the UV dose produces the mandatory crosslink between protein and target RNA, the risk of over-crosslinking increases, too. At 312 x 100 µJ/cm^2 UV light, the radioactive signal does not increase much further by additional irradiation, indicating that the optimal limit for one crosslink per RNA-protein complex is reached and additional bonds are only formed with binding partners that are already linked. The RNase amount should be optimised to produce RNA species of 50–150 nt lengths (Huppertz et al. 2014). Under-digestion will rise into a diffuse signal and can later lead to artefacts because truncation events will not be introduced randomly but occur due to endogenous ribonucleases which might have preferred sequence or structure specificity. Over-digestion, on the other hand, will give a sharp band that migrates close to the expected molecular weight of the protein. IR-L3 adapter ligation may therefore be hindered or the cDNAs produced will be too short for later mapping. Besides this conclusion, the blot clearly shows rather high background. Apparently, TALON™ either captures other proteins non-specifically or nucleic acids adhere to the beads. Neither washes under denaturing conditions with 6 M urea nor the more specific elution with imidazole could improve the background (data not shown).

To clear the high non-specific background, capture by the FLAG-epitope was taken into consideration as the epitope-antibody interaction is supposedly highly specific. Experiments were therefore performed replacing TALON™ by FLAG capture keeping the same conditions for UV crosslinking while adjusting RNase digestion slightly (**Figure 3-7 C**). The FLAG-antibody detaches from the epitope under the urea conditions which are essential to access the epitope tag. Therefore, the crosslinked samples are diluted back to lower urea before loading onto the anti-FLAG beads. During the iCLIP experiment shown here, the RNA was first 5'-end labelled with γ-^{33}P-ATP before the IR-L3 adapter was attached. Compared to the TALON™ capture, FLAG removes most background (**Figure 3-7**, radioblot). Strikingly, the dominant negative rPrp16-G378A binds RNA strongly even without crosslinking (**Figure 3-7 C** lane 5, 6). The blot visualising IR-L3 adapter via the infrared label, which should exactly mirror the radioblot when all RNA-protein complexes are ligated to the IR-L3 adapter, shows a lot of background, some of which might derive from adapter concatamers (**Figure 3-7 C**). Stringent washing conditions containing chaotropic agents are not compatible with the epitope-antibody interaction, which is clearly a disadvantage of FLAG over TALON™ capture.

Several tests combining both affinity purifications resulted in very low yield and similar background, and therefore were not considered as any improvement. Hereby, the protein was bound first to FLAG, eluted with either high urea or FLAG-peptide followed by TALON™ capture. Isolated tests, incubating magnetic beads under ligation conditions with IR-L3 adapter or IR-L3 adapter and RNA ligase I showed that the adapter preferentially adheres to TALON™ beads; this behaviour was significantly reduced when using FLAG beads (data not shown). In addition, IR-L3 adapters attached to proteins of lower molecular weight supposedly gives a stronger IR signal (Zarnegar et al. 2016), therefore background of lower molecular weight appears stronger than when visualised by radiolabelling (compare **Figure 3-7 C** lanes 1-6 and lanes 7-12). In the final *in vitro* spliceosome iCLIP protocol, an additional step was introduced to remove free IR-L3 adapter by enzymatic deadenylation and subsequent endonuclease digestion (communication with J. Ule and used e.g. in (Zubradt et al. 2017)). Despite the disadvantages discussed above, FLAG capture gave cleaner results and therefore all following experiments are carried through under FLAG capture.

Since many obstacles were encountered in establishing *in vitro* spliceosomal iCLIP for Prp16, SmB1 protein was introduced as a control for method development. The Ule lab routinely includes the SmB control using Sm B'/B/N antibody (Santa Cruz) when adapting conventional iCLIP to a novel protein in human cells. SmB is known to form crosslinks to spliceosomal snRNAs with high efficiency; it is an integral step-independent component of the spliceosome and its binding site is well described in the literature (Guthrie and Patterson 1988). To obtain a consistent control for *in vitro* spliceosomal iCLIP, a yeast strain expressing SmB-3xFLAG-His_8 was generated. Consequently, the main steps of the protocol could remain unchanged, merely the amount of UV and RNase I had to be adjusted. The optimal signal could be obtained with irradiation of 312 x 100 µJ/cm^2 UV light and 5 units RNase I (**Figure 3-7 D**). Other potential RNA-protein complexes appear at sizes around 65 kDa, where other proteins migrate which most likely crosslink to the same RNA and are released by this concentration of RNase I. To ensure correct analysis of the data, this part of the membrane was excluded when extracting the RNA-protein complexes.

3.4 RNA library preparation produces cDNA for subsequent high throughput sequencing

3.4.1 cDNA amplification correlates with crosslinked RNA amounts

Even though the infrared signal from the IR-L3 adapter did not always exactly mirror radiolabelled RNA, RNA was extracted and transcribed into a cDNA library. Most likely, 100 % ligation efficiency was never reached. Extraction of the RNA-RBP complex from the membrane was monitored by scintillation counting.

After reverse transcription and purification of the circularised cDNA, amplification was optimised and reached an optimum of 24 cycles in this representative experiment (**Figure 3-8 A**). Samples derived from SmB iCLIP show the strongest signal indicating its high crosslinking efficiency. In the same way, more cDNA is generated from the dominant negative mutant of Prp16 compared to the WT, underlining a higher RNA binding ability of the mutant protein. **Chapter 4.1.2** discusses the length of the pre-mRNA substrate as a crucial parameter for obtaining longer fragments of cDNA.

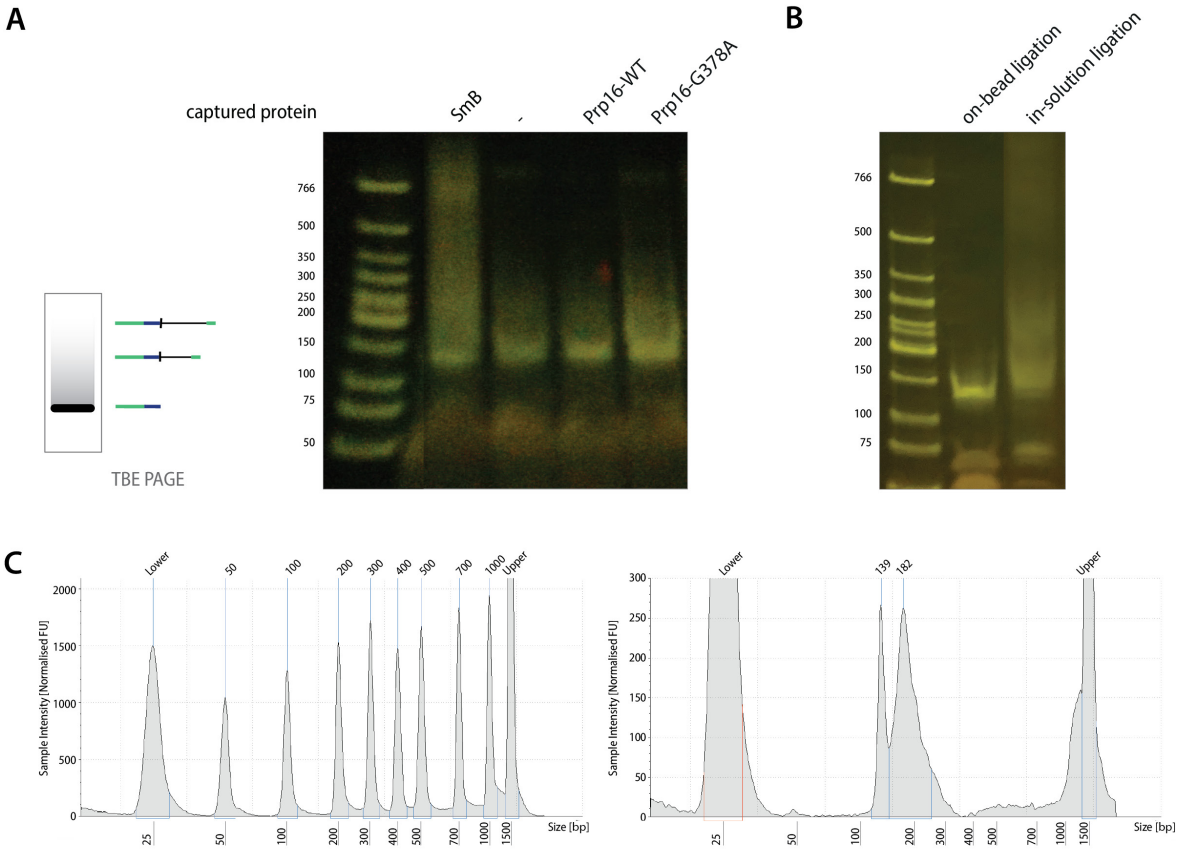


Figure 3-8: cDNA library preparation of *in vitro* spliceosome iCLIP. (**A**) 6 % TBE PAGE stained with SYBR green 1 shows cDNA library after multiplexing of non-irradiated and irradiated samples after *in vitro* spliceosome iCLIP of the protein as labelled. (**B**) The comparison between on-bead ligation and in-solution ligation of the IR-L3 adapter combined with the gel-free protocols shows a stronger signal for in-solution ligation. The same sample was split equally to pursue both strategies. The cDNA was amplified with 21 cycles each. (**C**) Size distribution of a representative multiplexed library of SmB, Prp16-WT and DN gives a peak around 139 nt which most likely refers to adapter and primer artefacts. The actual cDNA is presented by the peak around 182 nt reaching until about 400 nt. To determine DNA sizes and concentration with the Tapestation, the sample (right panel) is compared and aligned to a ladder (left panel).

Overall, most cDNA libraries in this thesis were amplified with 25 ± 2 cycles, which is consistent with conventional iCLIP for low abundance proteins. Despite introducing an adapter removal step, all libraries showed a strong band at ≤ 130 nt (compare **Figure 3-8 A**) which corresponds to artefacts derived from primers and IR-L3 adapters. Therefore, the libraries were gel purified at least once.

The P3/P5 Solexa primers and the barcode sequence combine to a length of 131 nt. All DNA fragments of 131 to ~400 nt length were extracted to retain all cDNA inserts which are larger than 15 nt. For quality control, the concentration and the length distribution of the final library was determined by Qubit and Tapestation, respectively. Libraries normally reached a yield between 200 and 1,000 fmol depending on the quality and the number of multiplexed samples. Often, the distribution trace showed two peaks, where the first, narrow one corresponds to IR-L3 adapter and primer artefacts and the second peak shows the desired shape of a library with various lengths of cDNA inserts (**Figure 3-8 C**).

3.4.2 Short versions of iCLIP generate higher yields of RNA

One of the main aims in iCLIP is the reduction of steps in the protocol to save time, retain more material and reduce the possibility of human error. While optimising *in vitro* spliceosome iCLIP, two shorter versions were tested. The main motivation to try the first shorter version derived from low capture of RNA requiring a relatively high amplification rate. In the so-called short-iCLIP protocol (personal communication with Flora Lee), the SDS-PAGE step is replaced by urea elution. The protein is released from the antibody-bound beads followed by a second immunoprecipitation from which the crosslinked RNA is directly released by proteinase K digestion. The protocol was adapted to the previously described *in vitro* spliceosome iCLIP conditions. Since the urea elution seemed inefficient, the crosslinked RNA was directly eluted by proteinase K digestion from the first FLAG capture. The amount of eluted RNA was determined higher by scintillation counting and the same amount of cDNA was generated with a lower number of amplification cycles. Therefore, the short-iCLIP protocol was an improvement superficially, however the protocol was not stringent enough to separate crosslinked RNA-protein from background (discussed in **chapter 3.5**, sequencing results).

In a second attempt, a gel-free version was examined (personal communication with Andrea Elser), in which the crosslinked RNA-RBP complex is first purified by a selective RNA column followed by protein binding beads (**7.4.4 Materials and methods**). This two-step purification was considered to retain only crosslinked RNA-RBP species. Although simplifying the experimental protocol, the main motivation for this trial was to reduce background in non-UV control conditions. In addition, this protocol offers the possibility of attaching the IR-L3 adapter after proteinase K digestion. By avoiding on-bead adapter ligation, steric hindrance of the crosslinked protein is reduced, which can in principle prevent efficient IR-L3 adapter ligation. Starting from a multiplexed cDNA library derived from the gel-free protocol, on-bead adapter ligation (left) and in-solution adapter ligation (right, ligation after elution from beads and proteinase K digest) were compared (**Figure 3-8 B**). While both libraries were amplified with 21 cycles, the resulting amount for in-solution ligation is much higher

indicating an increase in adapter ligation. Although the final cDNA library derived from the gel-free protocol showed the highest concentration with 4 pmol, a lower number of reads during sequencing were finally identified as unique cDNA reads, whilst the background was highly increased compared to the protocol including the SDS-PAGE (**Appendix folder**, session_1).

3.5 Next-generation sequencing data confirms Sm site

A novel bioinformatics pipeline was written in Snakemake (Köster and Rahmann 2012) for *in vitro* spliceosome iCLIP data based on the command-line interface iCount (Chakrabarti et al. 2017) to visualise crosslink positions on RNA (this pipeline was completely developed by Charlotte Capitanchik) (**Appendix folder**). First, a separate genome was generated that only contains the pre-mRNA substrate used for the *in vitro* splicing reaction and all spliceosomal snRNAs; from now on, this will be called 'substrate-ome' throughout the thesis. Reads were trimmed to remove IR-L3 adapter and barcode sequences. Subsequently, trimmed reads were mapped first to the substrate-ome and retained if they mapped uniquely and in the correct orientation (**Figure 3-9**). Unmapped reads were then mapped to the yeast genome. Due to the use of unique molecular identifiers (UMIs), PCR duplicates could then be collapsed based on the UMI and start position of the mapped reads, such that each crosslink represents one cDNA. The crosslink position corresponds to the -1 position of the sequence read. The number of truncation events per position is shown in histograms on the target RNA.

A first indication of the quality of the *in vitro* spliceosome iCLIP data is that control experiments without any tagged protein contain very few reads. Reads which map to the total yeast genome seem to truncate at random locations and fall below background levels and are therefore not shown (**Appendix folder**, session_2a and 2b for substrate-ome and genome mapped iCLIP data, respectively). The background level of reads in non-UV-irradiated control samples aligned to the substrate-ome is low for SmB, too (**Figure 3-9 A**). As expected, reads map to U2 and U5 snRNA, which are part of spliceosomal complex C. In addition, reads also align to U1 and U4 snRNA, which are part of earlier spliceosomal complexes present in the spliceosome preparation and not removed by additional purification of a second-step specific factor such as Prp16. Reads which map to U6 snRNA, which holds the LSm site rather than a Sm site, or the RNA substrate fall below background level. Denaturation with urea after crosslinking dismantles the spliceosome and therefore only RNA molecules that are in direct contact with SmB proteins are expected to give rise to cDNA reads. Less stringent conditions, however, could possibly produce cDNA reads which terminate at crosslink sites of auxiliary spliceosomal proteins through indirect contacts; this was shown in (Briese et al. 2018) applying conventional SmB iCLIP under mild conditions. The location of the peaks agrees with

the known Sm sites with the sequence element $PuAU_{4-6}GPu$ (Pu = purine) (Guthrie and Patterson 1988). On U2 snRNA, the Sm site locates close to the 5'-end at 110 nt. The Sm site on U5 snRNA is found at ~170 nt. A second peak maps to ~150 nt and allows speculations of an alternative binding site. In previous studies, the U5 Sm site showed to be very tolerant towards mutations (Jones and Guthrie 1990) and yeast was still viable when then the complete U5 Sm site was deleted (Dix et al. 1998; O'Keefe, Norman, and Newman 1996). The Sm site of U1 snRNA is found at 550 nt at its very 3'-end (Seipelt et al. 1999), whereas the site is located at 150 nt on U4 snRNA. All mapped regions locate slightly upstream of the $PuAU_{4-6}GPu$ sequence motif and the truncation events spread several bases (**Figure 3-9**). All these results validate *in vitro* spliceosome iCLIP.

The landscape for Prp16 binding in complex C shows a broad accumulation of truncations between the branchpoint adenosine (brA) and +5 nucleotides from the 3'SS (**Figure 3-10** and **Appendix folder**, session_3). Further, prominent peaks are located at the 5 'SS, ~+10 nt from the 5'SS, and at the brA. Hardly any reads were generated from experiments capturing the dominant negative mutant Prp16-G378A. In addition, reads derived from non-UV-irradiated control samples result in a similar landscape. Since control experiments without tagged proteins are very clean (compare lines 1 and 2 with lines 3-6), this background might be due to strong direct or indirect interaction of the helicase with its target RNA. The appearance of prominent peaks at the 5'SS and the brA will be discussed in **chapter 4**, which is fully devoted to Prp16 *in vitro* spliceosome iCLIP data interpretation. A negligibly small number of reads map to the snRNAs (**Figure 3-10**).

Short-iCLIP data reveals a landscape where most reads truncate at the 5'SS or the brA throughout control and non-control samples (**Appendix folder**, session_4). Since truncation events are also obtained from untagged protein samples, those reads are most likely Prp16 independent. Consequently, the additional steps including SDS-PAGE separation of the long protocol are necessary to reduce the amount of non-crosslinked RNA species present in the sample as well as the cleanliness of control samples.

In summary, this initial analysis of iCLIP data requires further improvement but certainly supports the applicability of *in vitro* spliceosome iCLIP for Prp16 helicase.

Figure 3-9: Mapping of SmB *in vitro* spliceosome iCLIP data to the substrate-ome. Crosslink events (-1 position of truncation) are presented as histograms aligned to the indicated RNA (U2 snRNA, U5 snRNA, U1 snRNA, U4 snRNA, U6 snRNA and the transcript 3xMS2_*UBC4*_AC_short). Boxed regions are enlarged on the right, showing the nucleotide sequence and the position of the first and last nucleotide. The Sm sequence element is highlighted in bold. Signal on U2, U5, U1, and U4 snRNAs locate to the Sm sites. Neither U6 nor the pre-mRNA substrate contains reads above background. The scale of the histograms (number of truncated reads/nucleotide) is given at the left corner in each lane. The data was generated applying the *in vitro* spliceosome iCLIP method that includes the SDS-PAGE separation step.

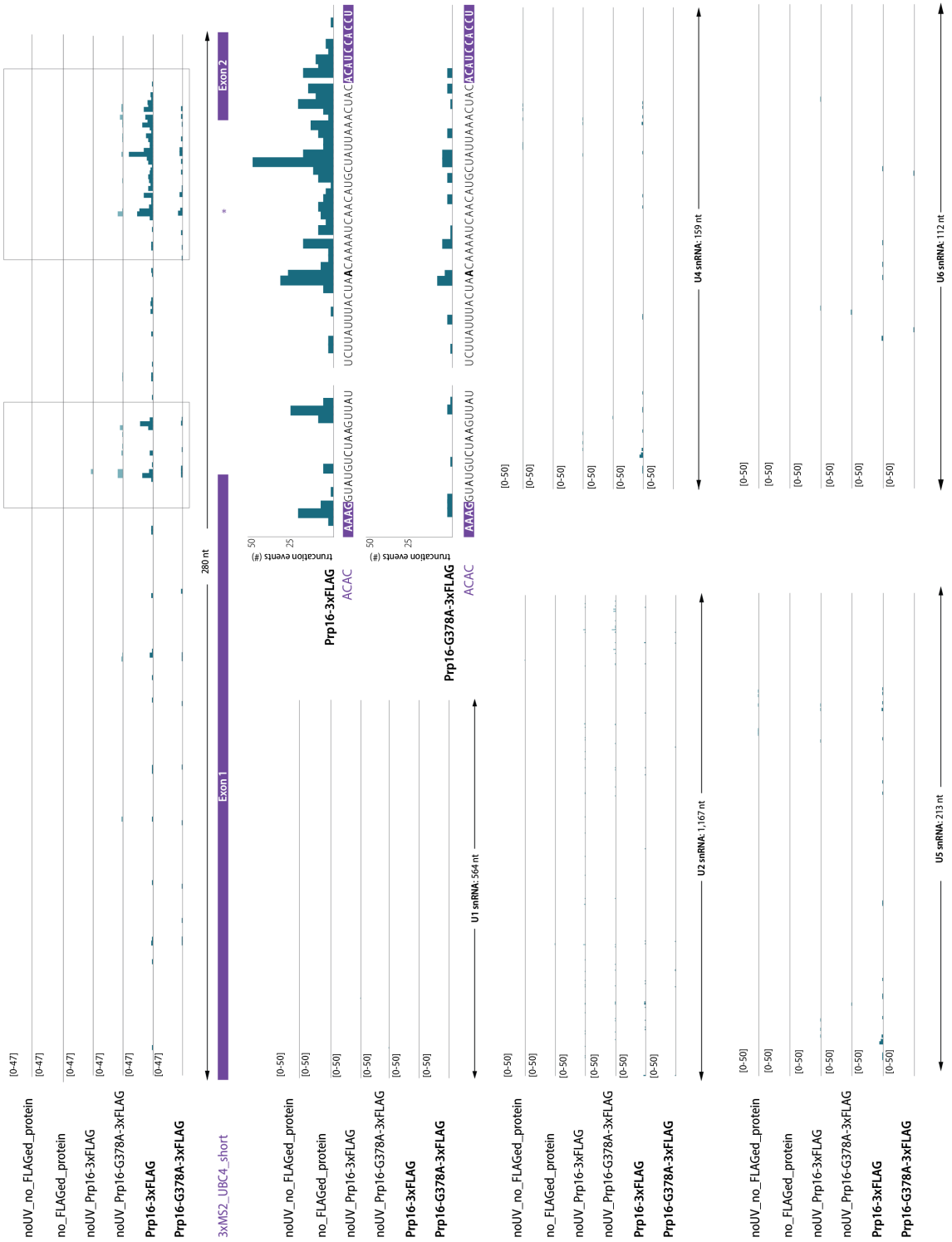


Figure 3-10: Mapping of Prp16-WT and Prp16-G378A *in vitro* spliceosome iCLIP data to the substrate-ome. Cross-link events (-1 position of truncation) are presented as histograms aligned to the indicated RNA (the transcript 3xMS2_*UBC4*_AC_short (exons are in purple, brA is represented by a *), U1 snRNA, U2 snRNA, U4 snRNA, U5 snRNA, U6 snRNA). Boxed regions are enlarged underneath, showing the nucleotide sequence and the position of brA (bold A), and the exons (purple). Reads accumulate between the brA and 3'SS. Sharp peaks locate to the 5'SS and the brA. Reads on the snRNAs are at background level. The scale of the histograms is indicated at the left corner or is given as the y-axes for enlarged parts. The data was generated applying the *in vitro* spliceosome iCLIP method that includes the SDS-PAGE separation step.

3.6 Helicases are suitable targets for iCLIP of defined spliceosomal states

In summary, this chapter describes the development of a complete pipeline, which catches the RNA-binding position of tagged proteins in spliceosomal complex C (summarised in **Figure 3-11**). The method was verified by SmB *in vitro* spliceosome iCLIP data, which generates the expected Sm site landscape on spliceosomal snRNAs.

The position of Prp16 could not be determined with similarly high confidence. The broad peak around +15 nt from the brA agrees with previously published data (McPheeters and Muhlenkamp 2003; Semlow et al. 2016), however, cleaner data would allow a more confident prediction. The number of cDNAs derived from rPrp16-G378A is much lower than from the WT protein, which is not in agreement with the stronger RNA binding activity of the mutant protein. Besides this, the proportion of cDNAs mapping to the substrate-ome is low as well (**Appendix Figure 8-2**). Binding of a protein close to the 3'-end of an RNA, naturally results in short cDNAs which is a limiting factor. Short cDNAs are more likely to have multiple possible mapping positions, making confident assignment more difficult. Therefore, cDNA reads shorter than 11 nt are excluded for data analysis in the bioinformatic pipeline. Since the landscape for Prp16 suggests a binding position close to the 3'-end generating a lot of short reads, this restriction might exclude a high number of reads lowering the overall signal at the binding site (see **Appendix, Figure 8-3**). In addition, a large number of truncation events happen at the intron-lariat junction. Both events are unique to *in vitro* spliceosomal iCLIP where the short RNA substrate is converted into intron-lariat intermediate. These problems are not present in the data for SmB *in vitro* spliceosome iCLIP since the target sites are located on linear RNA, most with a sufficient distance to the 3'-end.

Attempts at improvement such as increasing the length of the pre-mRNA substrate and including an enzymatic debranching step are discussed in **chapter 4**. The promising landscape for SmB and Prp16 *in vitro* iCLIP supports the potential for the method to be transferred onto other helicases in the spliceosome, as well as giving sufficiently accurate information to be used as a starting point for time-resolved *in vitro* spliceosome iCLIP.

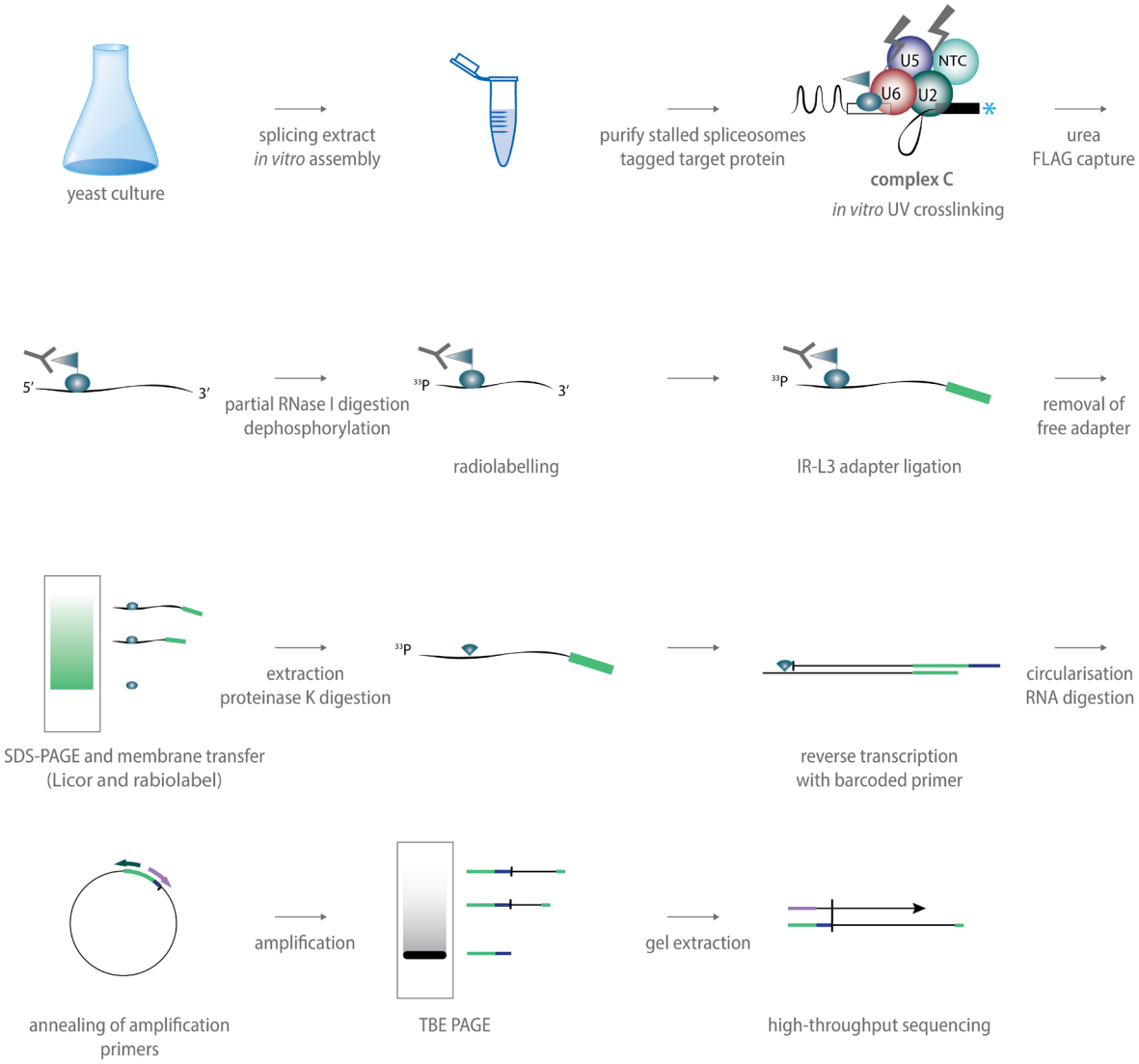


Figure 3-11: Workflow of *in vitro* spliceosomal iCLIP. Splicing extract is made from endogenously FLAG-tagged yeast strains or supplemented with recombinant protein expressing the FLAG-tag. Different strategies stall spliceosomes at various stages during the *in vitro* assembly, such as a substitution at the 3'SS of the pre-mRNA or dominant negative Prp16 for complex C. The spliceosomes are purified via the MS2-tag on the RNA and irradiated with 254 nm UV light to covalently link protein and RNA. Denaturation of the complex with urea makes the FLAG-tag accessible, which is necessary for Prp16. While capturing the protein of interest, controlled RNase I digestion fragments the RNA into suitable pieces for Illumina sequencing, the 5'-end is radiolabelled, and the 3'-end is dephosphorylated to allow IR-L3 adapter ligation. The covalently linked RNA-protein complexes are separated by SDS-PAGE, blotted onto nitrocellulose membrane, and extracted by proteinase K digestion. Reverse transcription truncates at the crosslink site occupied by the remaining protein peptide. Barcoded primers will identify reads from different experiments and the random sequence allows identification of reads derived from PCR duplicates later. The cDNA is circularised, and the RNA template digested. After further purification of the cDNA, it is amplified directly with P3 and P5 Solexa primers. The resulting library is further purified on a 6 % TBE PAGE to separate cDNA from free IR-L3 adapter and primer artefacts. After gel extraction and quantification, the libraries from different experiments are multiplexed and loaded onto the Illumina HiSeq 2500 system for single-read sequencing.

4 Prp16 binds around 15 nt downstream of the intron branch in order to remodel the catalytic core

In the previous chapter (**3**), *in vitro* spliceosome iCLIP was developed for the helicase Prp16 in spliceosomal complex C and the method was validated by determining the binding profile of SmB in spliceosomes at and before the first step, which reflects the expected Sm sites (Guthrie and Patterson 1988). The work described in this chapter aims at improving the quality of the *in vitro* spliceosome binding profile for Prp16 in order to analyse its RNA interactions with more confidence. In spliceosomal complex C, the DEAH-box helicase Prp16 is located at the periphery and binds to the intron-lariat intermediate somewhere downstream of the branch point (**Figure 4-1** and (Galej et al. 2016)). In order to remodel the spliceosome to a second step conformation, a rearrangement has to happen at the catalytic core. Prp16 is thought to initiate the remodelling by pulling on the transcript from its peripheral location (Semlow et al. 2016). The branch site sequence, which forms the branch helix together with U2 snRNA, moves out of its binding pocket in order to allow 3'SS docking in the catalytic centre (Galej et al. 2016; Fica et al. 2017; Wilkinson et al. 2017). Splicing factors of the first catalytic step, such as Cwc25 and Yju2 are released and binding sites for second step factors generated. Prp16 promotes, together with the second step factors, remodelling towards the exon-ligation conformation of the spliceosome (Schwer and Guthrie 1992a; Ohrt et al. 2013).

Through cryoEM studies, the location of Prp16 on the spliceosome has been determined, however, the path of the RNA transcript which would identify the binding site of Prp16 on the transcript could not be resolved in these flexible regions. The distance from brA to the RNA entry site of Prp16 was measured to be around 18 nucleotides for the *UBC4* transcript (Galej et al. 2016). Previous biochemical crosslinking studies, in which individual substrate RNA nucleotides were replaced by the photoreactive 4-thiouridine, located the Prp16 binding site within the last four nucleotides of the intron and the first 13 nucleotides of the second exon of the ACM_{13} transcript (McPheeters and Muhlenkamp 2003). ACM_{13} is an actin-derived pre-mRNA substrate, in which the distance between the branch site and the 3'SS is shortened to 13 nt, and commonly used in *in vitro* splicing assays. In a second study, spliceosomes were assembled on the *UBC4* transcript in which Prp16 crosslinked

to 4-thiouridine, which was introduced at the -8 position on the intron counted from the intron-exon boundary (Semlow et al. 2016).

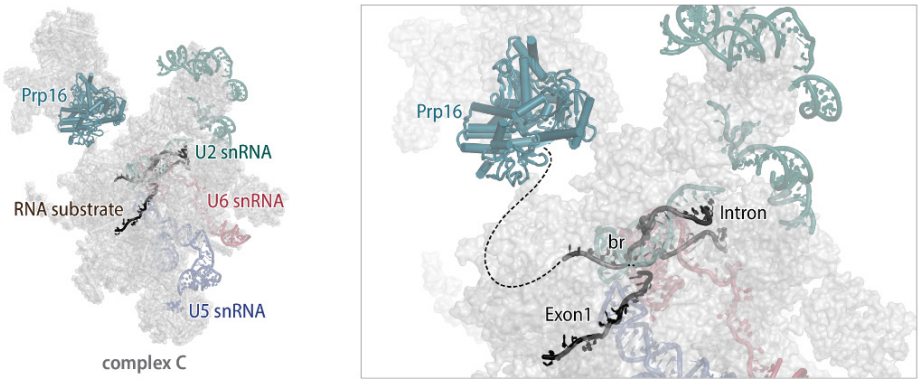


Figure 4-1: Structures of spliceosomal complex C highlighting the position of the helicase Prp16. Prp16 (dark cyan) locates at the periphery of the complex, whereas the catalytic core is in the centre of the particle (Galej et al. 2016). The RNA substrate which consists of exon 1 and the intron-lariat-exon2-intermediate (location of exon 1 and intron are indicated), the snRNAs, and the branch (br) are highlighted. Left: overview of the particle; Right: zoom-in of the helicase region and the target RNA; dashed lines represent some of the unmodeled RNA parts which should be connected.

As described in **chapter 3.5**, Prp16 *in vitro* spliceosome iCLIP reveals a binding profile in which most truncation events accumulate between the positions of -22 and +5 nucleotides of the second exon with the highest peak at -11 on the 3xMS2_*UBC4*_AC substrate (*UBC4*_AC: 3'SS alteration to AC). This region includes the position found by 4-thiouridine replacement in (Semlow et al. 2016); the direct comparison is possible since both experiments were based on the *UBC4* transcript.

Compared to SmB *in vitro* spliceosome iCLIP, which reveals precise locations of the Sm sites, the interpretation of Prp16 *in vitro* iCLIP data is less straightforward. First, the proportion of reads mapped to the substrate-ome is low (**Appendix Figure 8-2**, < 5% for Prp16), while a relatively high number of truncation events is obtained from non-UV irradiated samples. In this chapter, a strategy using an extended version of the transcript will be discussed as means for improvement. Second, prominent peaks at the 5'SS and the branch point adenosine can in principle derive from either peptide- or RNA-induced terminations. The latter are caused when the reverse transcription stops at the 2'-5' phosphodiester bond of the intron-lariat intermediate, which could obscure real cross-link events and lead to misinterpretation of the data. In this chapter, debranching of the intron-lariat intermediate is explored for *in vitro* spliceosome iCLIP on catalytic step 1 spliceosomes.

Finally, the chapter aims at giving a general picture of where the natural binding site of Prp16 is located before the helicase starts translocating to promote remodelling towards the second step conformation. Three different datasets are compared to analyse Prp16 in complex C: (1) the WT protein stalled in the complex using a transcript with a 3'SS mutation, (2) the dominant negative protein Prp16-G3678A on the same transcript, and (3) Prp16-G378A which stalls the spliceosome

through its ATPase deficiency in the context of an unmodified RNA transcript. The Prp16-G378A carries an amino acid substitution in helicase motif I (Walker A) where the consensus sequence AxxGxGKT, which forms a phosphate binding loop required for the interaction with Mg^{2+} and ATP, is altered to AxxGx**A**KT. The conserved residue G in GKT is supposed to maintain a flexible conformation of this loop (Walker et al. 1982).
In addition, the binding profiles of Prp16 on two different substrates are compared to identify what determines the start location for Prp16 translocation.

4.1 The binding profile of Prp16 WT and mutant protein is identical

4.1.1 Truncation events co-localise for Prp16 WT and mutant protein

In *in vitro* spliceosome iCLIP, the intrinsic crosslinking activity of the RNA enables one to locate the binding site of a protein on its unperturbed target RNA within its natural context, the spliceosome. Here, we aim at identifying the location of Prp16 before translocation starts and remodelling is initiated or in other words the RNA binding site for Prp16 in spliceosomal complex C. Both the alteration of the 3'SS of the RNA transcript (Vijayraghavan et al. 1986) and ATPase deficient dominant negative mutant proteins of Prp16 (Schneider, Hotz, and Schwer 2002) have shown to stall splicing at the complex C stage. Since it was never confirmed to my knowledge, that the two strategies lead to the exact same spliceosomal state, it is important to determine Prp16 binding profiles with both. This will allow us to identify the location of Prp16 before remodelling with high confidence. Therefore, this subchapter describes the generation of *in vitro* spliceosome iCLIP data for rPrp16-G378A on *UBC4*_WT substrate for comparison with the results obtained in the previous chapter (**Figure 3-10**, data on *UBC4*_AC).

To set up *in vitro* spliceosome iCLIP for complex C solely stalled by the dominant negative Prp16-G378A, the same concentration of rPrp16-G378A as in the context of 3xMS2_*UBC4*_AC was used here. This concentration was sufficient to inhibit the second step of splicing comparable to stalling with the *UBC4*_AC substrate (**Figure 4-2 A** and **Figure 3-3 B**). The purification yield of the assembled spliceosomes on the 3xMS2_*UBC4*_WT transcript was comparable to that obtained using 3xMS2_*UBC4*_AC. The proportion of pre-mRNA and intron-lariat-intermediate spliceosomes after the purification was also similar (compare **Figure 4-2 A** to **Figure 3-3 B**).

Further sample preparation followed the pipeline developed in **chapter 3**. The obtained binding profile for Prp16-G378A on 3xMS2_*UBC4*_WT transcript is aligned with the two datasets for Prp16 on the 3xMS2_*UBC4*_AC transcript for comparison (**Figure 4-3**, for Prp14 *UBC4*_AC data also compare **Figure 3-10**; **Appendix folder**, session_3 and 5). The overall peak pattern derived from the

three experiments is identical: the major peaks locate at the 5'SS, +12 nt on the intron, and the brA, while an accumulation of truncation events locates between the brA and the beginning of the second exon.

Potential concern in the interpretation of the data is the background, which is obtained from samples without UV irradiation, in which peaks resemble the pattern found in UV-irradiated samples with lower intensity (**Figure 4-3**, all three experimental conditions).

Further, most truncation events map to the last 40 nucleotides of the transcript. Short reads are likely to be excluded from mapping by not passing the 11 nt length threshold or are lost during the cDNA purification when separating them from similarly sized primer and IR-L3 adapter artefacts.

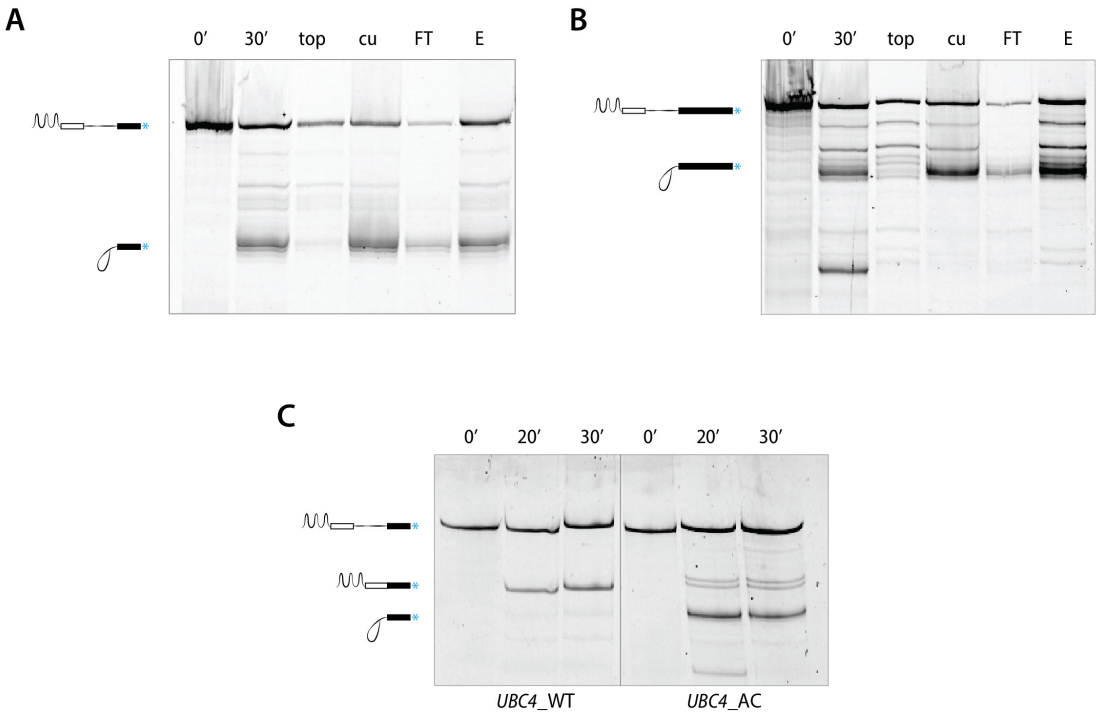


Figure 4-2: Spliceosomes are stalled at the complex C stage by dominant negative rPrp16-G378A. (**A**) Spliceosome assembly on 3xMS2_*UBC4*_WT_short transcript in rPrp16-G378A supplemented splicing extract, followed by glycerol cushion and amylose purification. (**B**) Spliceosome assembly on 3xMS2_*UBC4*_WT_long transcript in rPrp16-G378A supplemented splicing extract, followed by glycerol cushion and amylose purification. (**A** and **B**) RNA species are visualised by the Cy5-label on a 10 % denaturing PAGE: before splicing (0') after 30 min incubation (30'), top (top) and cushion (cu) after centrifugation through the glycerol cushion, flow-through (FT) and elution (E) of the amylose capture. The band with the fastest migration in (**B**) appearing at 30' corresponds to debranched lariat-intermediate. (**A**) and (**B**) were run side-by-side on the same gel but split for the figure, meaning the migration is directly comparable. For both splicing substrates, spliced RNA would migrate between pre-mRNA and intron-lariat intermediate on the 10 % denaturing PAGE. (**C**) This 10 % denaturing PAGE shows a comparison of splicing reactions of 3xMS2_*UBC4*_WT_short and 3xMS2_*UBC4*_AC_short in absence of the dominant negative rPrp16-G378A to confirm that the band indicated as intron-lariat intermediate in (**A**) corresponds to intron-lariat intermediate. Originally, the lanes were run on the gel in a different order than shown here.

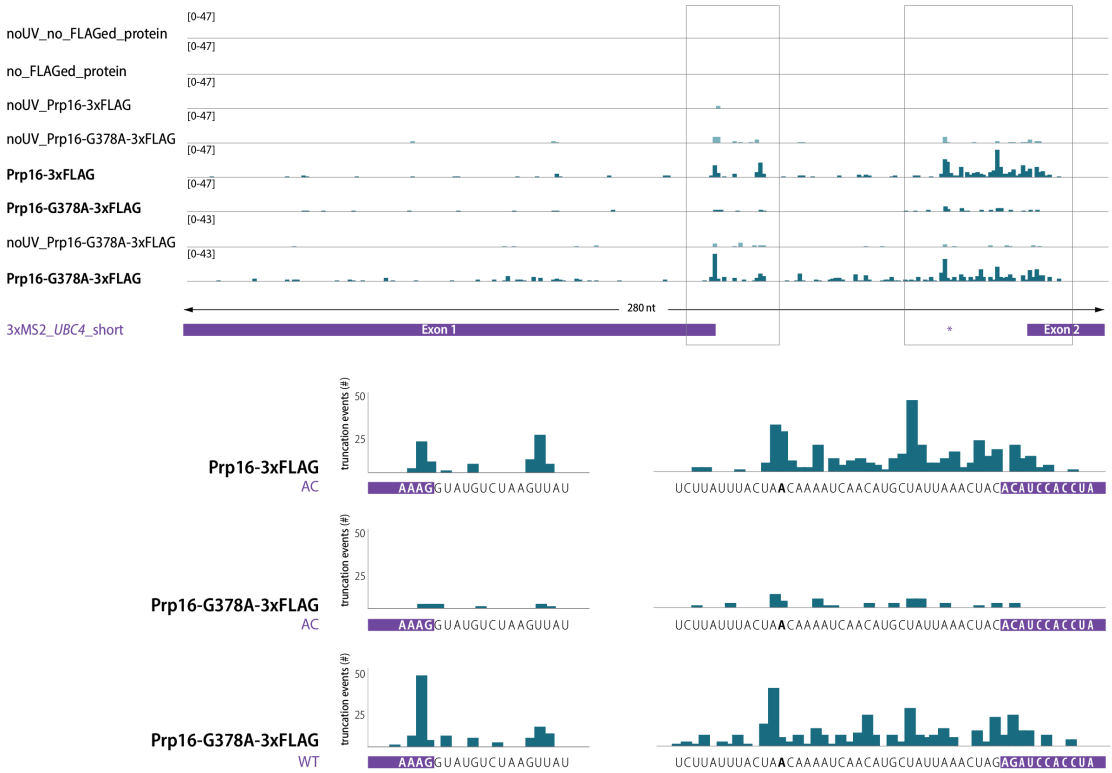


Figure 4-3: Prp16-WT and Prp16-DN produce the same binding profile in *in vitro* spliceosome iCLIP. Crosslink events (-1 position of truncation) are presented as histograms aligned to the 3xMS2_*UBC4*_AC_short or 3xMS2_*UBC4*_WT_short transcript as indicated in the enlarged parts. Exons are depicted in purple; the brA is marked with a *. Boxed regions are enlarged underneath, giving the nucleotide sequence, while highlighting the position of brA (bold A) and the exons (purple). The scale of the histograms is given in the left corner or as the y-axis. Read truncations accumulate between the brA and 3'SS. Further peaks locate to the 5'SS, +12 nt on the intron and to the brA. Reads mapped to the snRNAs are at background level and therefore not shown henceforth for Prp16 *in vitro* spliceosome iCLIP data (included in **Appendix folder**, session_3 and 5).

4.1.2 Extension of the transcript on the second exon increases number of reads

To increase the probability for reads to map, the length of the 25 nt long second exon of *UBC4* was increased by 36 nt (named 3xMS2_*UBC4*_long). With a complete exon length of 61 nt, the required minimal 11 nt for unique mapping should be exceeded by far.

Samples for *in vitro* spliceosome iCLIP were prepared as previously described for 3xMS2_*UBC4*_short (**Figure 4-2 B** for spliceosome purification).

The proportion of reads which map to the substrate-ome increased substantially (**Appendix Figure 8-2**, **Appendix folder** session_6 and 7). The histogram height representing the number of truncation events increased up to 14-fold (**Figure 4-4** compare dark cyan and red bars). The position and prominence of the other sharp peaks remained unchanged.

Overall, the binding profile for rPrp16-G378A gives histograms of lower intensity compared to Prp16-WT (**Figure 4-3**, **Figure 4-4** and **Figure 4-7**). Since the proportion of reads mapping to the yeast genome, however, does not differ between the two proteins, rPrp16-G378A should not carry

pre-bound RNA from its purification into the iCLIP procedure but its crosslinking efficiency might simply vary (**Appendix Figure 8-2**).

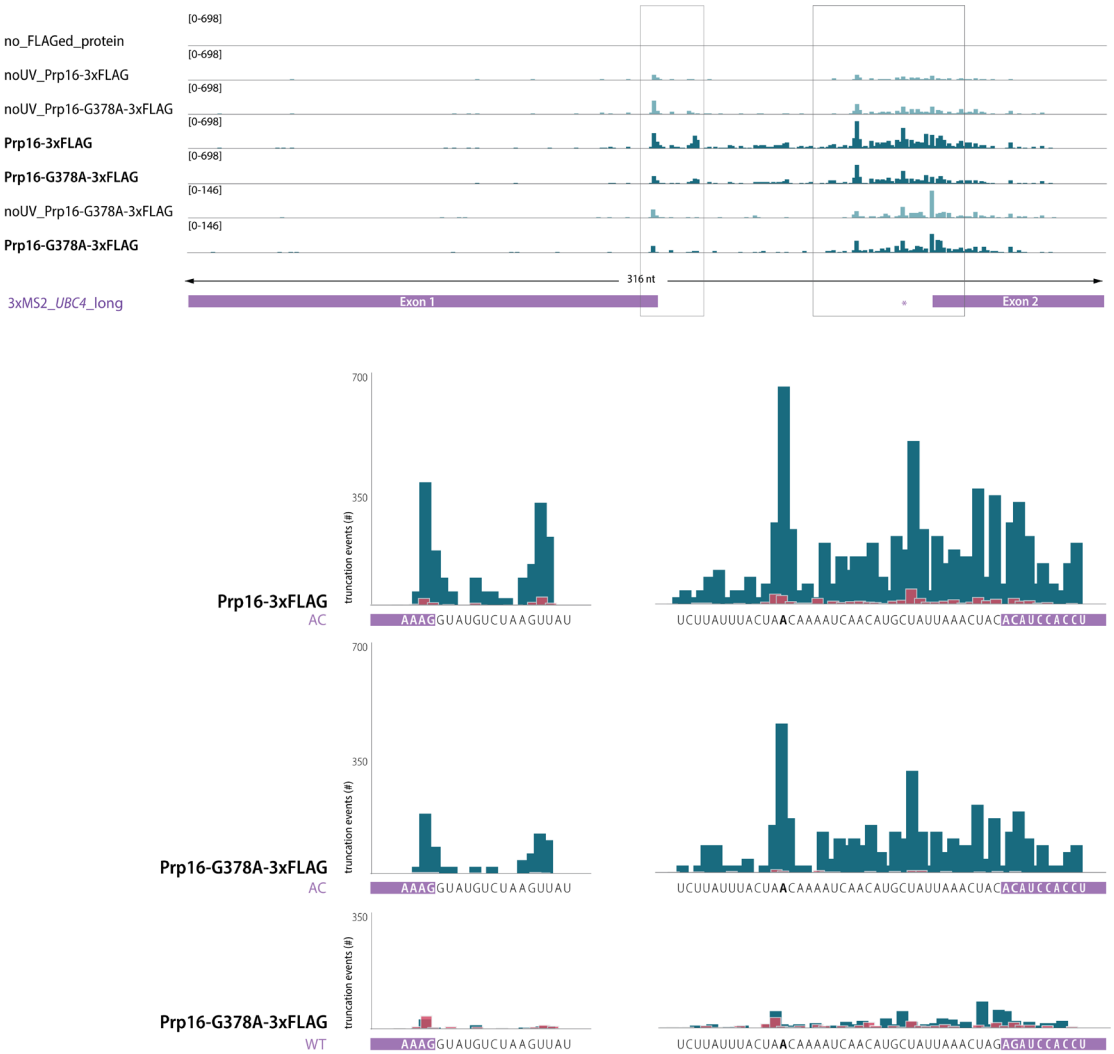


Figure 4-4: Increase of the transcript length increases the number of uniquely mapped reads on the substrate-ome. Crosslink events (-1 position of truncation) are presented as histograms aligned to the transcript 3xMS2_*UBC4*_AC_long or 3xMS2_*UBC4*_WT_long as indicated in the enlarged parts. Exons are highlighted in light purple; the position of the brA is marked with a *. The scale of the histograms is given in the left corner or as the y-axis. Boxed regions are enlarged underneath, showing the nucleotide sequence, the position of the brA (bold A), and the exons (purple). Reads accumulate between the brA and 3'SS. Further peaks locate to the 5'SS, +12 nt on the intron and to the brA. Dark cyan histograms are derived from the transcript 3xMS2_*UBC4*_long and compared in scale to the transcript 3xMS2_*UBC4*_short, which is presented in red (compare **Figure 4-3**).

The first peak on the *UBC4* transcript directly corresponds to the 5'SS and a second peak corresponds to the brA (**Figure 4-4**). These two peaks also dominate the pattern of control samples (light cyan histograms). The peaks are likely derived from reverse transcription stopping at the intron-lariat branch rather than at a Prp16-derived covalently liked peptide. When reverse transcription starts downstream of the branch but is not terminated by a crosslinked peptide of Prp16, the tran-

scriptase will be stopped in most cases by the obstacle caused from the additional 2'-5' phosphodiester bond at the brA. SuperScript III stops at this junction in 85 % of cases (Chen et al. 2018). When the RT is transcribing downstream of the 5'SS towards the junction, it is even less likely that the enzyme will continue upstream of the brA by crossing the 2'-5' phosphodiester bond. The termination rate for SuperScript III was determined to be 99 % (Chen et al. 2018). Reads starting from upstream of the brA on the intron-lariat intermediate will therefore terminate at the 5'SS. The generation of these reads in unavoidable, since all RNA fragments which include the intron-lariat junction will expose two 3'-ends after RNase digestion and therefore offer a second starting point for reverse transcription. Further, in contrast to positions of genuine crosslink events, which are transcribed by RT with an accuracy of > 90 %, a high rate of mutations is found here. The brA itself is identified as an adenine in only 62 % of reads and is read as cytosine in > 30 % of reads (**Figure 4-5 A**, right). The nucleotide, which would be upstream of the 5'SS is identified as a mix of adenine, guanosine, and uridine (**Figure 4-5 A**, left for **x**|GU). The mutation rate leads to the conclusion that these two sharp peaks correspond to the 5'SS and brA, respectively. By monitoring terminations caused by lariat-junctions, (Chen et al. 2018) successfully predicts branch points in yeast introns by spliceosome profiling. Similarly, conventional SmB iCLIP performed under native conditions was shown to generate cDNA reads of intron-lariat splicing intermediates that allowed the prediction of branch points (Briese et al. 2018).

Assuming the peak at the 5'SS is not a crosslink-site, it can be used as a reference point to compare the amount of signal in the region downstream of the brA between UV-irradiated and non-irradiated samples. More truncation events are clearly acquired in the presence of UV (**Figure 4-4**, rows 1-6 for experiments of this subchapter). I conclude that truncation events must have therefore been also induced by UV dependent protein crosslinks. Nevertheless, obtaining a similar binding pattern around the crosslink site in control samples is quite likely when a protein binds strongly to its target RNA. In the previous chapter, both Prp16-WT and mutant protein were shown to bind RNA after the stringent iCLIP procedure in the absence of UV irradiation (compare **Figure 3-7 C**). The bound RNA fragments, which have an average length of 30 nt (**Appendix Figure 8-3**), align close to the protein binding site and therefore generate a crosslink-like profile. Similar biases in iCLIP data derived from inappropriate read length is discussed in (Haberman et al. 2017).

Another strong peak is observed around position +12 nt in the intron (AG|UU, +10 to +13 nt). The hypothesis of an alternative splice site is weakened by the correct assignment of a guanine nucleobase just upstream in more than 90 % of the reads; the signal can most likely be considered as a genuine crosslink site. When comparing this crosslink site with the 5'SS as reference point, about three times less crosslinks are obtained for rPrp16-G378A on the 3xMS2_*UBC4*_WT transcript than

of Prp16 and rPrp16-G3678A on the 3xMS2_*UBC4*_AC transcript. To my knowledge, an interaction site for Prp16 at this position has not been reported yet. Since the sample subjected to *in vitro* spliceosome iCLIP contained spliceosomes before and at the first catalytic step of splicing purified further only through Prp16, the crosslink could in principle derive from stages before or at complex C.

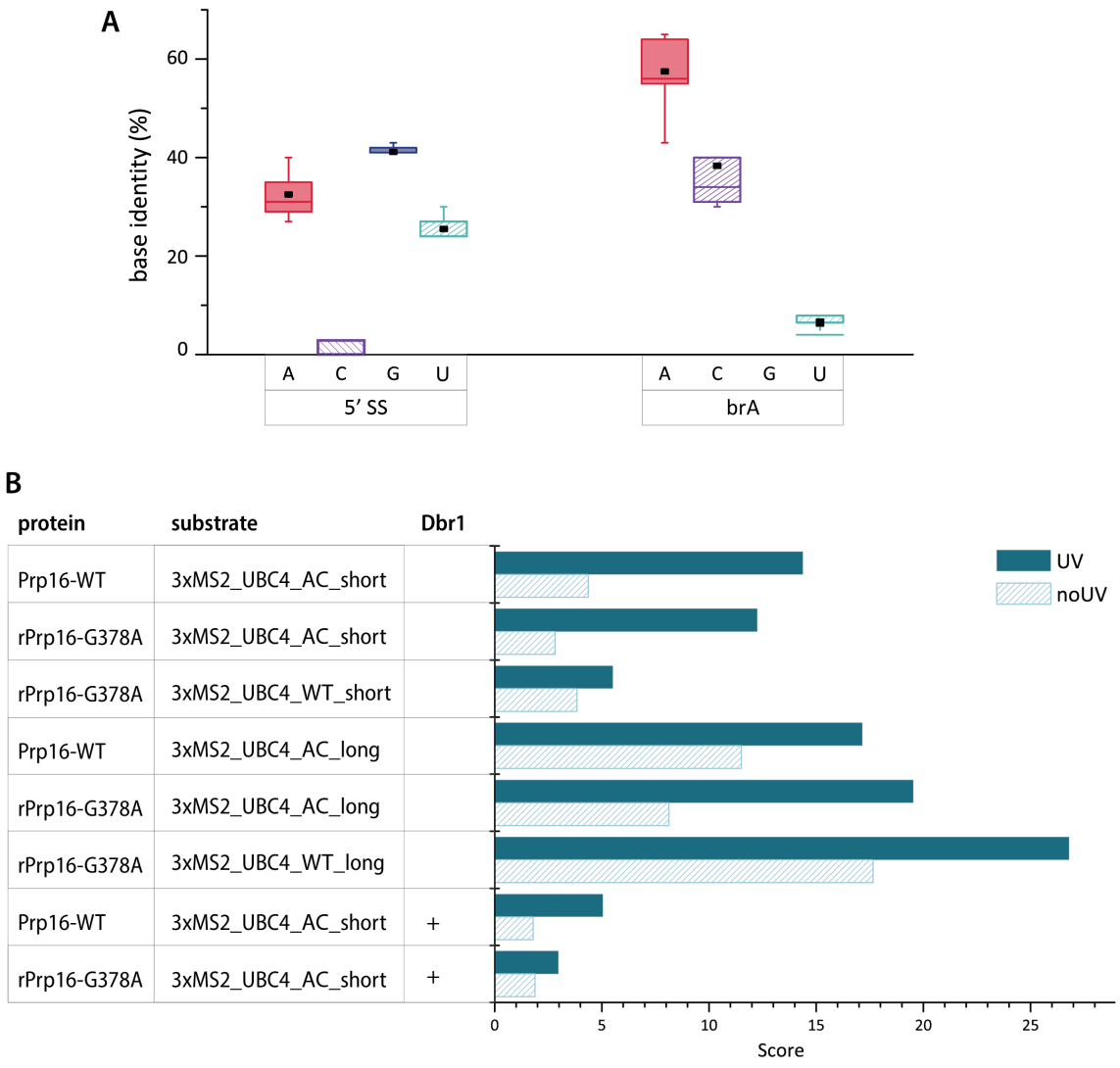


Figure 4-5: Quantitative comparison of peaks at the 5'SS, brA, and the region downstream of the brA. (**A**) Nucleobases at the 5'SS (left) and the brA (right) are not read correctly by RT. At the 5'SS the position **x**|GU (wherein | indicates the splice site) was analysed. In the event of a non-spliced product a G and in the event of a continuing read in the intron-lariat an A would follow. At the brA, the read accuracy for the branch point adenosine was analysed directly. Numbers include the experiments shown in **Figure 4-4** (excluding 'no_FLAGed_protein' control) and are taken from all reads, including transversion reads, to be compared in a box plot. Black squares give the mean. (**B**) Data derived from UV-irradiated samples show a higher proportion of reads in the region downstream of the brA in comparison to at the 5'SS on *UBC4* substrates. Scores are calculated as follows: the sum of truncation events of position 234 to 279 nt (corresponds to the region downstream of the brA) divided by the number of truncation events at position 162 (corresponds to the 5'SS).

The remaining truncation events predict the Prp16 binding site to around +15 nt from the brA (**Figure 4-3** and **Figure 4-4**). The region, in which these crosslinks locate to, spans at least 15 nucleotides;

a distance which is longer than one would expect for the occluded site of a DEAH-box helicase (Tauchert et al. 2017). Uridines have a higher tendency than other nucleotides to be crosslinked to protein by 254 nm UV light (Sugimoto et al. 2012), which might explain the more prominent peaks at uridine positions compared to other nucleotides in this area. As discussed earlier, a similar position was found, when a photoreactive 4-thiouridine was introduced to the +18 position in *UBC4* (Semlow et al. 2016).

Further, the area in which most crosslink events are mapped is consistent between the three tested conditions (**Figure 4-3** and **Figure 4-4**). Neither the location of the mutant nor the WT protein is shifted by the mutation at the 3'SS. Most likely, Prp16-WT protein would also crosslink at this exact position in the 'natural' complex C without interfering through a splice-site mutation induced blockage. This finding is important as it determines the start location and is a first reference point for Prp16 time-resolved translocation experiments in the future.

4.2 Debranching reduces truncation events at brA

4.2.1 Dbr1 acts only on the branch in the absence of the spliceosome

Debranching the intron-lariat intermediate before reverse transcription was considered as a possibility to reduce peaks at the branched position. Both the human and yeast RNA lariat debranching enzyme (Dbr1) was shown to hydrolyse the 2'-5' phosphodiester bond of the intron-lariat intermediate, while leaving a 2'-hydroxyl group at the branch attachment site and a 5'-phosphate end at the 5' SS (Ooi et al. 2001).

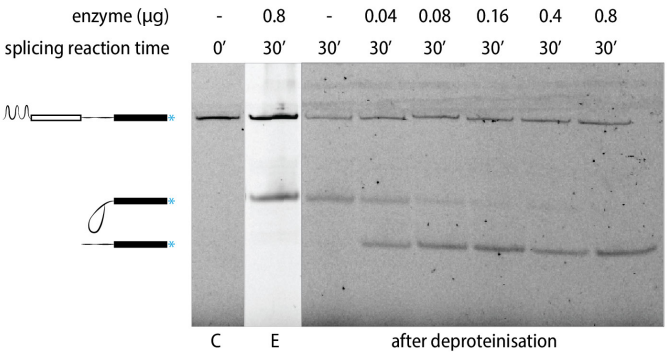


Figure 4-6: Dbr1 debranches protein-free 2'-5'-phosphodiester-bond-linked lariats. The control shows pre-mRNA before splicing (lane C). After splicing, the reaction was either directly incubated with 0.8 µg Dbr1 (lane E), or the RNA was first extracted and then incubated with the indicated amounts of Dbr1 (lanes: after deproteinization). Each single experiment was set up with 50 fmol Cy5-labelled pre-mRNA. 0.4 µg enzyme efficiently open the junction of 50 fmol RNA, which equals a concentration of 8 µg enzyme per pmol RNA. The three parts of the gel (C, E, after deproteinization) are derived from the same gel, on which they were originally run in a different order.

In a first step, commercial human Dbr1 from recombinant source was tested for its debranching activity. The branched intron-lariat intermediate was generated through *in vitro* splicing reactions

with the 3xMS2_*UBC4*_AC_short transcript stalling splicing after the first catalytic step. Subsequently, the whole reaction was either incubated with Dbr1 (first condition) or the obtained RNA was deproteinised and purified before incubation with Dbr1 (second condition). In the first condition, the intron lariat intermediate is not affected by the debranching enzyme (**Figure 4-6**, lane E), indicating that the spliceosome hinders access of the substrate to the enzyme. As a result, the enzymatic debranching step cannot be inserted into the *in vitro* spliceosome iCLIP procedure before denaturation of the spliceosome (compare **Figure 3-11** for the experimental pipeline). In the second condition, in which the RNA was isolated prior to treatment with Dbr1 (Chapman and Boeke 1991), the enzyme efficiently resolves the 2'-5' phosphodiester bond at a concentration of 8 ng enzyme per pmol RNA (**Figure 4-6**). Urea, unfortunately, impaired the enzymatic activity even at low concentrations (tested by C. Norman), so that debranching with Dbr1 was introduced as the first step during FLAG capture in the *in vitro* spliceosome iCLIP procedure. Extensive, prior washing with the reaction buffer should ensure total removal of residual urea.

Due to low RNA levels at this stage, the activity of Dbr1 cannot be monitored in *in vitro* spliceosome iCLIP. In addition, a possibility remains that Prp16, which crosslinks close to the brA, hinders substrate accessibility for Dbr1 and thereby interferes with debranching.

4.2.2 Dbr1 reduces truncation events at the brA

To analyse the effect of Dbr1 on *in vitro* spliceosome iCLIP, both Prp16 and rPrp16-G378A containing spliceosomes assembled on 3xMS2_*UBC4*_AC substrate were subjected to the *in vitro* spliceosome iCLIP procedure including the additional debranching step during FLAG capture. The peak at the 5'SS is unlikely to disappear through debranching, as most truncation events will still happen at the natural end of the transcript.

Treatment with Dbr1 reduced the prominence of the peak located to the brA by more than threefold when comparing to the height of the 5'SS peak (**Figure 4-7**, both Prp16-WT and Prp16-DN; **Appendix folder**, session_8). Similarly, the read coverage beyond the brA (upstream of the brA) increased in all experimental conditions through Dbr1 treatment, which supports that a number of intron-lariat junctions were successfully debranched during the *in vitro* spliceosome iCLIP procedure (**Figure 4-8**).

In principle, a complete elimination of the peak would be desirable. I am still working on optimising the conditions for the debranching enzyme. The debranching step could also be moved to a later stage in the protocol, when Prp16 is broken down to a small peptide at the attachment site. Further improvement of Dbr1 reaction conditions combined with the use of the 3xMS2_*UBC4*_long transcript, will hopefully lead to clearer results.

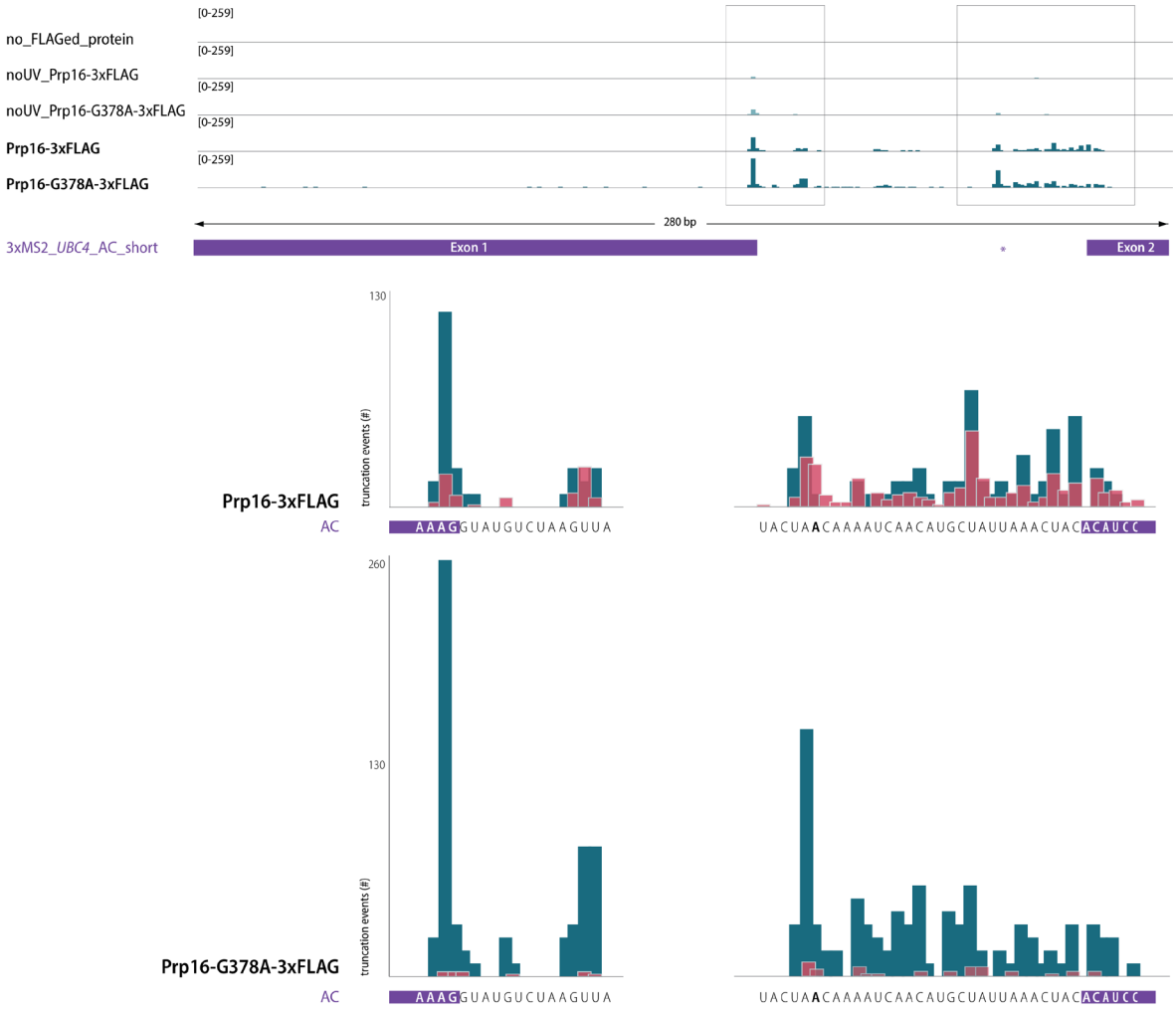


Figure 4-7: Dbr1 reduced the truncation events at the brA. Crosslink events (-1 position of truncation) are presented as histograms aligned to the transcript 3xMS2_*UBC4*_AC_short. Exons are highlighted in purple; the location of the brA is indicated with a *. The scale of the histograms is given in the left corner or as the y-axis. Boxed regions are enlarged underneath, showing the nucleotide sequence and the position of brA (bold A), and the exons (purple). Reads accumulate at similar positions with and without the debranching step. Dark cyan histograms are derived from *in vitro* spliceosome iCLIP preparation including the debranching step and compared to data without Dbr1 treatment **Figure 4-3** (red) in scale. The total histogram height for the mutant protein was increased.

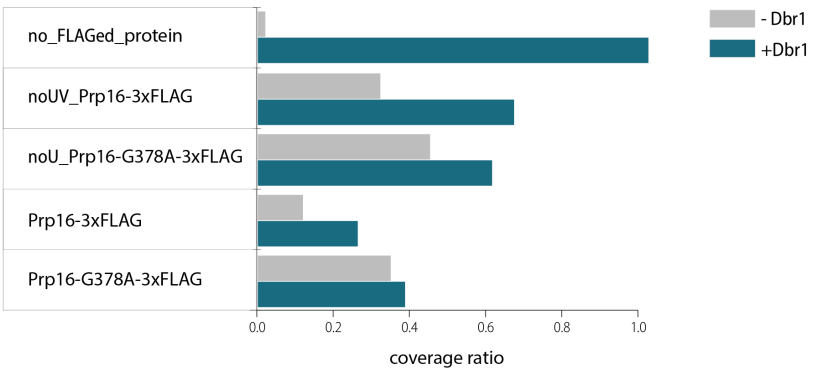


Figure 4-8: Read coverage increases for the region upstream of the brA through Dbr1 treatment. Bars show the ratio between the coverage upstream and downstream of the brA. The coverage is defined as number of aligned reads per nucleotide in the respective region on the pre-mRNA substrate. Experiments depicted in **Figure 4-3** (-Dbr1, grey) and **Figure 4-7** (+Dbr1, cyan) were compared; all conditions show an increase in the coverage ratio, when Dbr1 treatment was included in the procedure.

4.3 The binding profile of Prp16 WT and mutant is identical on *ACT1*

4.3.1 *ACT1* can be used for *in vitro* spliceosome iCLIP sample preparation

Up to this point, all *in vitro* spliceosome iCLIP data was generated from spliceosomes assembled on *UBC4* transcript. *UBC4* is popular in biochemical and structural studies of pre-mRNA splicing due to its high splicing efficiency and its short intronic sequence. As already shown for the length of the second exon, short stretches of RNA are not necessarily beneficial for crosslinking and sequencing approaches. The likeliness of one RNA molecule being both crosslinked to a protein peptide while being part of the lariat junction will decrease with longer distances between the recognition sequences (5'SS, brA, 3'SS).

At the same time, general conclusions about Prp16 binding profile can only be gained through the comparison of different transcripts. As a second transcript, I have chosen *ACT1*, which codes for yeast actin, and is used throughout the literature for splicing studies both *in vitro* and *in vivo* (Domdey et al. 1984; Frank and Guthrie 1992; Tardiff and Rosbash 2006). The precise transcript used for *in vitro* spliceosome iCLIP carries three MS2 stem loops at the 5'-end, followed by 25 nt of the first exon, the intronic *ACT1* sequence with 309 nt, and 164 nt of the second exon. All distances between the recognition sites are extended compared to the previously used *UBC4* pre-mRNA transcript.

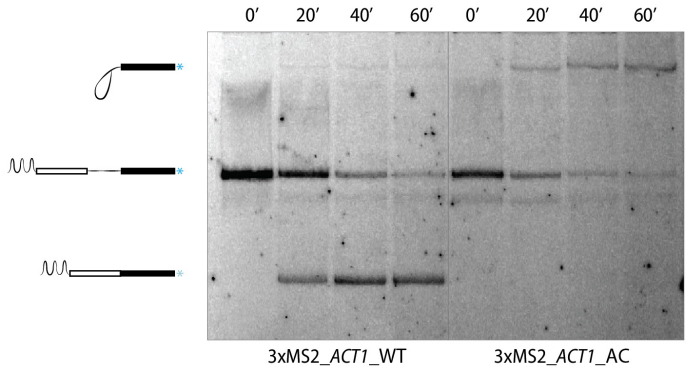


Figure 4-9: *In vitro* splicing of *ACT1* transcripts. 3xMS2_*ACT1*_WT (left) and 3xMS2_*ACT1*_AC (right) produce mRNA and intron-lariat intermediate after 60 min, respectively. The Cy5-labelled RNA species were separated on a 7 % gel.

Both 3xMS2_*ACT1*_WT and 3xMS2_*ACT1*_AC were generated and tested in *in vitro* splicing. To reach a similar splicing efficiency as for *UBC4*, the incubation time had to be doubled (**Figure 4-9**). *In vitro* iCLIP experiments were set up in the same way as carried out previously with *UBC4*. The amounts of rPrp16-G378A protein were as before, too. First, spliceosomes were assembled on 3xMS2_*ACT1*_*AC* and purified via amylose capture of the MS2-tag (as for 3xMS2_*UBC4*_AC in **Figure 3-3**). In a second step, the UV dose and RNase I levels were titrated to the same dose and a 20-times higher amount of enzyme as before, respectively to obtain optimally crosslinked RNA-protein

complexes; suitable conditions were evaluated as before (shown for *UBC4* in **Figure 3-7**). Enzymatic debranching was included during FLAG capture as described in **chapter 4.2**.

4.3.2 Binding profiles for *ACT1* show reduced background

Mapping of the data to the *ACT1* substrate shows that the overall histogram score is low compared to *UBC4* (**Figure 4-10**, score 5; **Appendix folder**, session_9 and 10). On the other hand, the pattern between control samples and UV irradiated samples is more distinct for *in vitro* spliceosome iCLIP on the *ACT1* transcript.

Both peaks at the 5'SS and the brA correlate with incorrectly interpreted nucleobases indicating RNA-mediated truncation events once more (similar nucleobase distribution as in **Figure 4-5 A**). The peak at the brA is prominent, most likely debranching was not completely efficient. Most reads terminate at locations between the brA and the beginning of the second exon, which is consistent throughout the three conditions and also in good agreement with the data derived from *UBC4* substrate. It is necessary to perform replicates of the experiment while increasing the number of uniquely mapped reads to be more confident about the crosslinking sites.

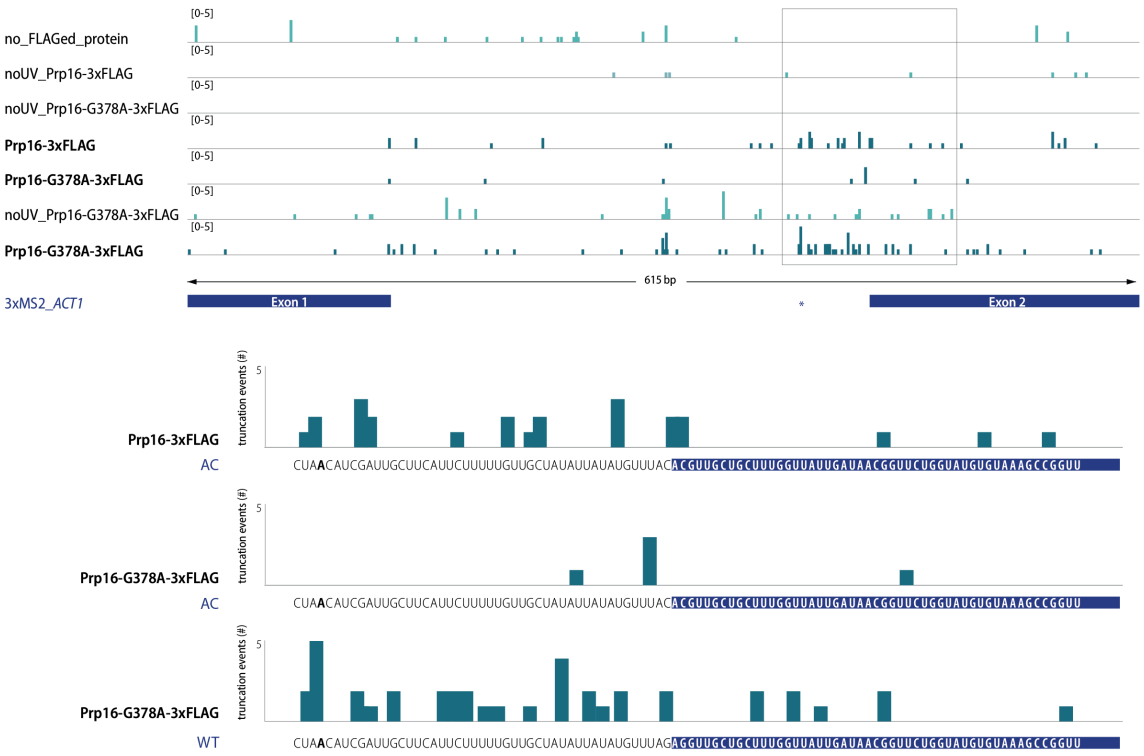


Figure 4-10: Prp16-WT and Prp16-G378A produce the same binding profile on *ACT1* transcript. Crosslink events (-1 position of truncation) are presented as histograms aligned to the transcript 3xMS2_*ACT1*_AC or 3xMS2_*ACT1*_WT as indicated in the enlarged parts. Exons are highlighted in navy; the location for the brA is indicated with a *. The scale of the histograms is given in the left corner or as the y-axis. Boxed regions are enlarged underneath, showing the nucleotide sequence and the position of brA (bold A), and the exons (navy). Read truncations accumulate between the brA and 3'SS (around +15 nt from the brA). Sharp peaks locate to the brA.

4.4 Prp16 binds similar locations on *UBC4* and *ACT1*

Taken together, this chapter has demonstrated how several issues arising specifically for Prp16 crosslinking and sequencing in the context of a first catalytic step spliceosome can impede data interpretation and how they can be resolved.

First, *in vitro* spliceosome iCLIP is restricted by the RNA length downstream of the crosslink site. By increasing the length of the second exon in the *UBC4* pre-mRNA, which doubled the distance from the major crosslink site to the 3'-end, the number of mapped truncation events could be increased by up to 14-fold.

Second, two sharp peaks in the *in vitro* spliceosome iCLIP data on the *UBC4* transcript could be interpreted as RNA-mediated truncation events. The 5'SS showed a strong peak throughout the experiments on *UBC4*, even in the absence of a pulldown protein. Similarly, many truncation events repeatedly mapped to the brA which could be reduced by including an enzymatic debranching step in the *in vitro* spliceosome iCLIP procedure.

Overall, the same binding profile is produced for Prp16 in complex C from three different conditions: (1) Prp16-WT on AC substrate and the ATPase-deficient helicase Prp16-G378A on both (2) AC and (3) WT substrate. Since the mutant protein is non-functional in splicing and has a residual ATPase activity of only 2 % (Schneider, Hotz, and Schwer 2002), which should hinder its helicase activity, the binding location in the spliceosomes stalled at complex C should correspond to the starting point for Prp16 translocation before remodelling.

In vitro spliceosome iCLIP data derived from *ACT1* transcript made it possible to draw sequence- and substrate-independent conclusions or Prp16 binding. Most truncation events accumulate around +15 nt from the branch on the intron of *UBC4*. Similarly, the major interaction site for Prp16 was found to be about 15 nucleotides downstream of the brA on *ACT1* (**Figure 4-11**). This suggests that Prp16's binding site on the transcript in complex C is determined through a molecular ruler mechanism. An obvious measure could be the spliceosome itself providing the binding surface for Prp16 in exactly this distance as implied by the three-dimensional structure of the particle (compare **Figure 4-1**). Like other members of the DEAH-box helicase family, Prp16 is expected to act in a sequence independent fashion and find its target for instance through protein-protein interactions and the spliceosome architecture. To prove this hypothesis, the distance between the brA and the determined binding site could be artificially increased in either *UBC4* or *ACT1* and analysed by *in vitro* spliceosome iCLIP.

The determination of Prp16's position in complex C supports the model of Prp16 acting from a distance and is the first essential step towards time-resolved *in vitro* spliceosome iCLIP experiments in the future. CryoEM studies have confirmed that the spliceosome core stays almost unchanged

through the C to C* transformation supporting the model that Prp16 will act from a distance while mediating this remodelling event.

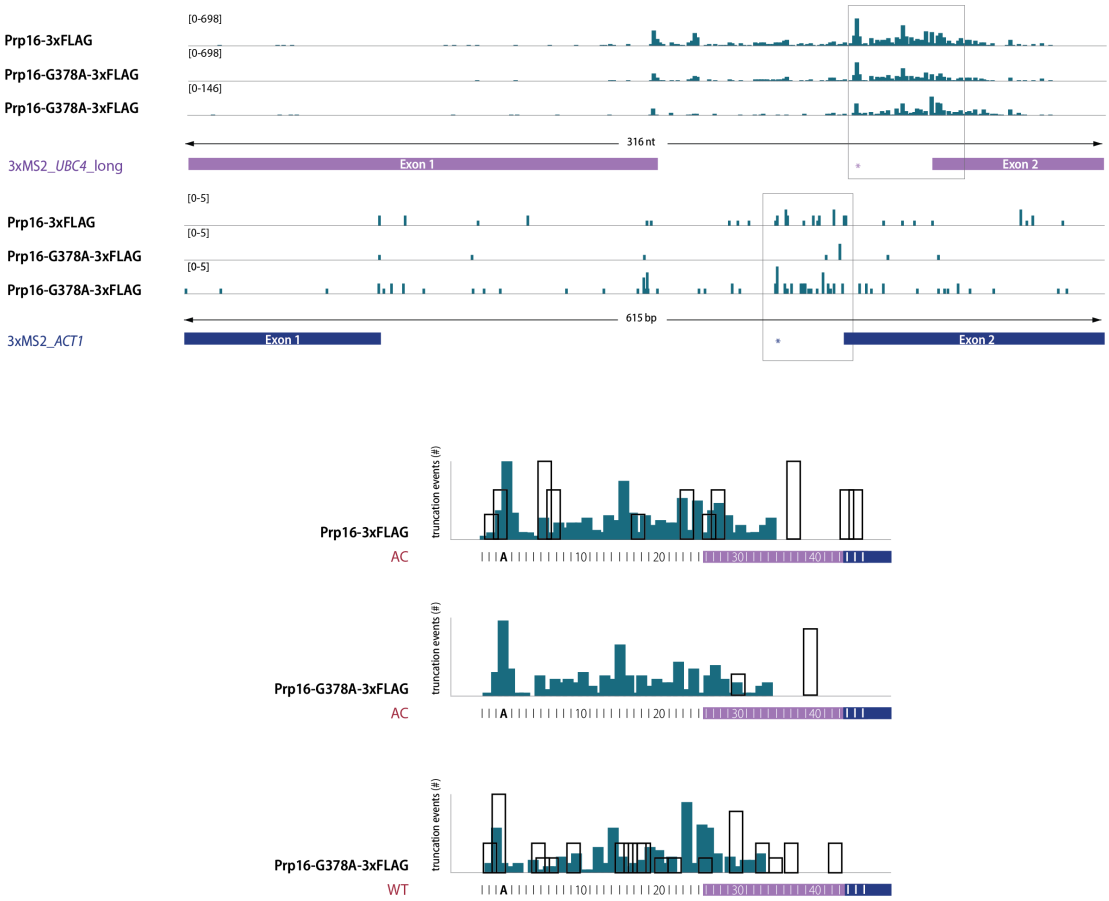


Figure 4-11: Comparison for Prp16 *in vitro* spliceosome iCLIP on *UBC4* and *ACT1*. Binding profiles of Prp16 on the transcript 3xMS2_*UBC4*_long (first three lanes) and on the transcript 3xMS2_*ACT1* (second three lanes) derived from Prp16-WT on AC substrate, Prp16-G378A on AC, and Prp16-G378A on WT substrate (indicated in the enlarged parts). Control lanes are depicted in **Figure 4-4** and **Figure 4-10**. Exons are highlighted as bars (purple for 3xMS2_*UBC4*_long and navy for 3xMS2_*ACT1*), while the location of the brA is marked with a *. The scale of the histograms is given in the left corner. The boxed regions are enlarged at the bottom (not to scale) and aligned to the brA. The profiles derived from *UBC4* (dark cyan) and *ACT1* (black boxes) are overlaid for each condition. Nucleotides are counted from the brA (bold A) and every 10th is numbered; the second exon is coloured in purple and navy for *UBC4* and *ACT1*, respectively.

5 Setting up *in vitro* spliceosome iCLIP to investigate additional spliceosomal helicases

In **chapter 3**, I describe the development and validation of the method '*in vitro* spliceosome iCLIP'. This chapter describes pilot experiments how the method can be modified to investigate additional spliceosomal helicases.

Whenever a novel target protein is analysed, the protein-specific steps in the iCLIP protocol have to be adjusted anew. In addition, the spliceosomal complex in which the helicase acts has to be stalled specifically and purified to provide a controllable system. Obtaining a homogeneous and tight stall is especially important if *in vitro* spliceosomal iCLIP will be employed to observe time-resolved helicase activity. The **subchapters 5.3** and **5.4** discuss a possible set-up to monitor the translocation of Brr2 and Prp16 on their target RNA over time and give a first impression of its feasibility.

Prp22 helicase is essential for mRNA release after ligation (Schwer and Gross 1998). Prp22 and Prp16, which both belong to the DEAH-box helicase family, are structurally very similar. In spliceosomal complex C, C* and P, they occupy the same position at the periphery of the spliceosome in a mutually exclusive fashion (Galej et al. 2016; Fica et al. 2017; Wilkinson et al. 2017). Prp22 was proposed to act from a distance pulling the ligated exons out of the spliceosomal core (Semlow et al. 2016), where the mRNA is held by base-pairing with U5 snRNA (Newman and Norman 1992; Sontheimer and Steitz 1993; Aronova et al. 2007) (compare **Figure 5-1 B**). The exact binding site for Prp22 in complex C* could not be precisely determined by cryoEM, as the density for the RNA substrate emerging from the core towards Prp22 is very weak (Fica et al. 2017) (**Figure 5-1 A**). In complex P, however, the local resolution is sufficient to locate the occluded site of Prp22 to approximately nucleotides 14 to 21 on the second exon of *UBC4* (Wilkinson et al. 2017) (**Figure 5-1 B**). The location is in agreement with the position found by 4-thiouridine crosslinking when mRNA release was blocked with a dominant negative Prp22 mutant (Schwer 2008). The first rearrangement mediated by Prp22, which requires additional recruitment of the second step factors Slu7 and Prp18, leads to the formation of the mRNA and was predicted to be ATP-independent (Schwer and Gross

1998). To release the mRNA from the spliceosome, however, Prp22 requires ATP (Schwer and Gross 1998; Wagner et al. 1998).

My intention to determine the binding site for Prp22 on its RNA target arose before the structure of complex P was solved. Nevertheless, a clear determination of its binding site in a stalled complex will become essential, when Prp22 translocation will eventually be analysed by time-resolved *in vitro* spliceosome iCLIP. The strategy which revealed the binding profile for Prp16 in complex C (compare **chapter 3.6**), can equally be applied to Prp22 stalled in either complex C* or P.

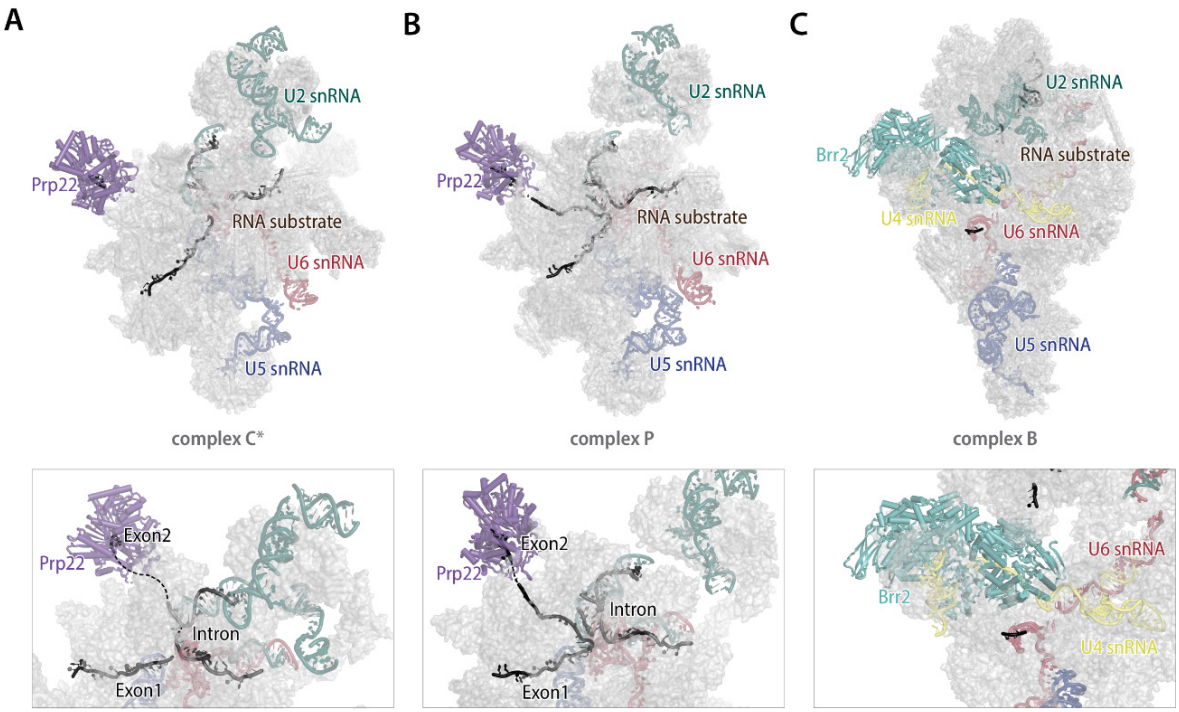


Figure 5-1: Structures of spliceosomal complexes highlighting helicase position and target. (**A**) In spliceosomal complex C*, Prp22 (purple) locates at the periphery of the complex, whereas the catalytic core is in the centre of the particle (Fica et al. 2017). The RNA substrate (exon 1, intron-lariat-exon2-intermediate) and the snRNAs are highlighted. (**B**) Spliceosomal complex P shows Prp22 at a similar position at the periphery of the complex (Wilkinson et al. 2017). The ligated exons, the intron and the snRNAs are highlighted. (**C**) Spliceosomal complex B shows Brr2 helicase (light cyan) loaded onto U4 snRNA which is base-paired with U6 snRNA (Hahn et al. 2012; Nguyen et al. 2016; Plaschka, Lin, and Nagai 2017). Parts of the pre-mRNA and the snRNAs are highlighted. (**A-C**) Top panels: overview of the particle; Bottom panels: zoom-in of the helicase region and the target RNA; dashed lines represent some connected RNA parts which are disordered in the structure.

The Ski2-like helicase Brr2 is already pre-loaded onto its target RNA, U4 snRNA, when it joins the spliceosome. Brr2 mediates the complex B to B^{act} transition by unwinding the U4/U6 snRNA duplex (Laggerbauer, Achsel, and Luhrmann 1998; Raghunathan and Guthrie 1998; Kim and Rossi 1999) Biochemical analysis and structural approaches (Nielsen and Staley 2012; Plaschka, Lin, and Nagai 2017) have mapped this pre-unwinding location of Brr2 to about nucleotides 70–90 of U4 snRNA (**Figure 5-1 C**). To date, unwinding of Brr2 was characterised in minimal recombinant systems or in tri-snRNP, neither of which reflect the natural context of the spliceosome (Small et al. 2006;

Mozaffari-Jovin et al. 2013; Henning et al. 2017). Brr2 was also proposed to be involved in the second step of splicing but its exact role remains unclear (Hahn et al. 2012). Following the path of Brr2 in a time-resolved manner would be an excellent approach to understand in which order specific RNA structures are actively remodelled by Brr2. Other structures might passively adopt a new conformation as a result of Brr2 helicase activity elsewhere in the complex. The complex B to B^{act} transition is a major rearrangement in which at least 24 proteins and the U4 snRNA are released while 22 other proteins join the complex (Plaschka, Lin, and Nagai 2017; Will and Luhrmann 2011). The factors required for the transition can be provided directly from splicing extract, avoiding the need for recombinant protein production. To make this transition targetable for *in vitro* spliceosome iCLIP, the protocol developed in **chapter 3** has to be modified. **Chapter 5.3** gives a first impression how Brr2 can be crosslinked to RNA in spliceosomal complex B and how the complex can be chased towards the catalytic stages.

Following the path of Prp16 on the intron-lariat intermediate requires a quite different strategy. A pre-steady state kinetic iCLIP experiment would be suitable to describe the translocation of Prp16 along its RNA target. Prp16-depleted spliceosomes were shown before to stall at the complex C stage and can be advanced to the next stage by addition of rPrp16 and ATP. This set-up was used for single molecule experiments that provided insights into the RNA substrate conformation (Semlow et al. 2016) and can serve as a starting point for time-resolved *in vitro* spliceosome iCLIP for Prp16. In **chapter 5.4** preliminary experiments were designed to deplete Prp16 from splicing extract and to determine the RNA binding ability of Prp16 under different conditions.

5.1 *In vitro* spliceosome iCLIP setup for Prp22 in C* complex

5.1.1 Purification of spliceosomal complex C*

Spliceosomal complex C* corresponds to the state after Prp16 remodelling but before exon ligation. Mutation of the pre-mRNA substrate at the 3'SS prevents exon ligation (**Figure 5-2 A**). To set up *in vitro* spliceosome iCLIP, a deoxynucleotide was introduced into 3xMS2_*UBC4*_short to replace the G| (where | indicates the 3'SS); the same substitution was used to solve the structure of complex C* (Fica et al. 2017). The following |AG was substituted with an |AC to block use of this cryptic splice site. Similar to blocking the second step of splicing with an AC|AC substitution for complex C, this stall can be combined with either WT or dominant negative Prp22. A WT substrate combined with a dominant negative version of Prp22, however, does not lead to complex C*. The spliceosome would efficiently catalyse the ligation step but fail to release the ligated exons, yielding spliceosomal complex P (Wilkinson et al. 2017). This method is used to obtain complex P in **chapter 5.2.** Firstly, a yeast strain expressing Prp22-3xFLAG was produced by tagging the genomic copy of *PRP22.*

We also produced recombinant Prp22-K512A-3xFLAG in yeast and purified it in two steps via CBP- and polyhistidine-tags (**7.3.1 Materials and methods**, **Figure 5-2 B**). The lysine is part of helicase motif I (Walker A) where the triad G**K**T forms the phosphate binding loop which normally coordinates the nucleotide. The dominant negative Prp22-K512A protein was shown to have a highly reduced ATPase activity and limited activity to release ligated exons (Schneider, Hotz, and Schwer 2002).

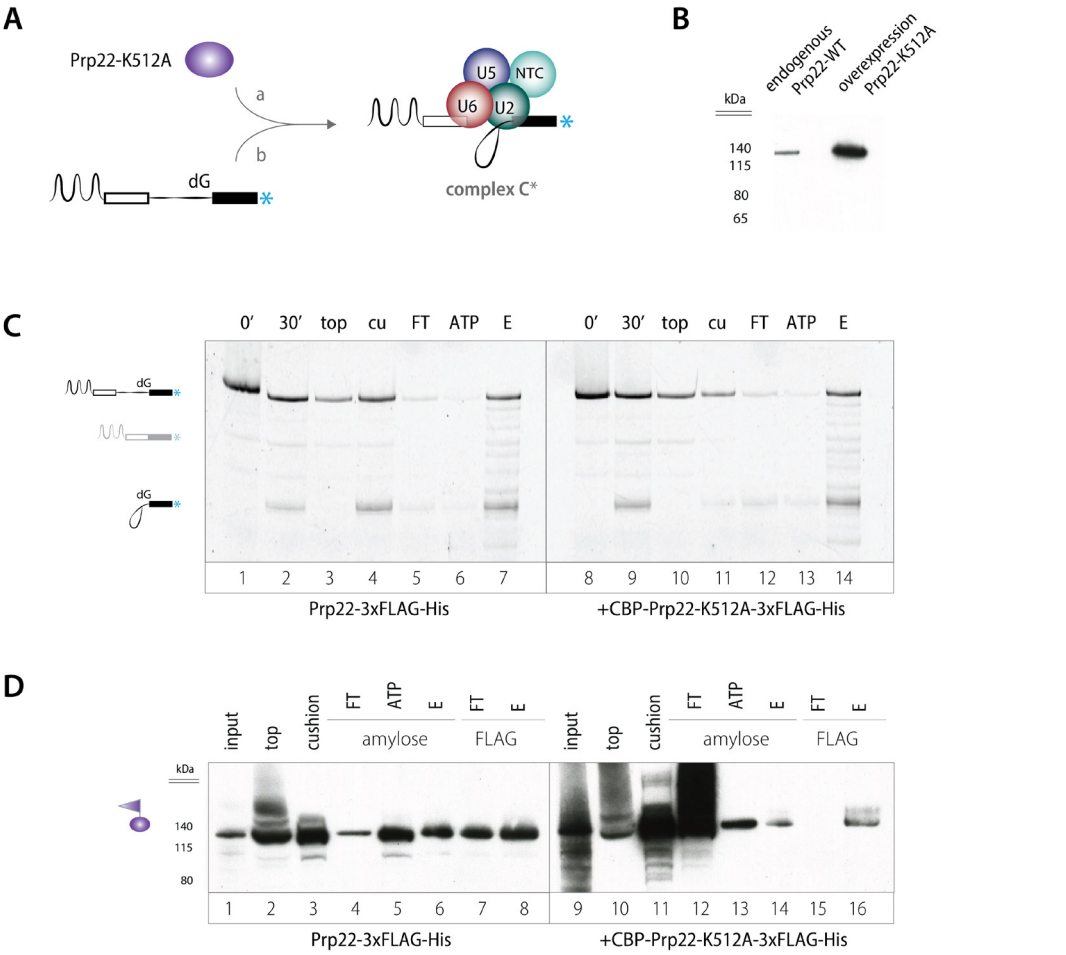


Figure 5-2: Purification of spliceosomal complex C*. (**A**) Substitution with deoxyguanosine at the 3'SS leads to the stall of complex C* which can either be combined with Prp22-WT or a dominant negative version of Prp22 (here Prp22-K512A). For visualisation during the purification procedure, the pre-mRNA substrate is labelled with Cy5 (indicated with a blue *). (**B**) Western blot probing with anti-FLAG antibody for endogenous Prp22-3xFLAG in whole yeast extract (left) or for rPrp22-K512A-3xFLAG overexpressed in yeast extract (right: cells were induced with galactose and harvested at a similar OD compared to the cells used for the left lane). A 10-fold excess of total protein was loaded on the left. (**C**) Purification overview of the splicing substrate separated on a 10 % denaturing PAGE. After splicing, purification via amylose capture of the substrate produces a mixture of spliceosomes of the pre-catalytic and first catalytic step. The faint band which migrates with a mobility similar to the spliced RNA is most likely a product spliced at a cryptic splice site (see grey spliced RNA indication on the left of the gel). Lanes are as follows in both (**C**) and (**D**): before splicing (0', only in (**C**)) after 30 min splicing reaction (30', lane input is a 5-fold dilution), top (top) and bottom (cu) of the glycerol cushion, flow-through (FT) as well as ATP incubation (ATP) on amylose and elution (E) thereof, flow-through (FT) and elution (E) of the FLAG capture. (**D**) Western blot probing for Prp22 throughout the spliceosome preparation followed by FLAG capture, either derived from splicing extract expressing endogenous Prp22-3xFLAG (left) or splicing extract supplemented with rPrp22-K512A-3xFLAG (right). The image contains two separately run western blots (left and right).

After assembly in an *in vitro* splicing reaction, spliceosomes were firstly captured via the RNA substrate on amylose beads using the MS2-MBP system. To stimulate the rearrangement of spliceosomes from complex C to the C* stage, which is Prp16 dependent, an additional incubation with ATP and magnesium was introduced during the capture as in (Fica et al. 2017). Despite Prp16 being bound to the spliceosome, Prp22 can presumably be loosely associated with the complex. Some Prp22, however, is lost during this incubation step (**Figure 5-2 D**, lane 5 and 13). A mixture of spliceosomes bound to pre-mRNA and intron-lariat-intermediate is eluted from amylose beads, some of which complex C*, as indicated by the presence of Prp22 (**Figure 5-2 C** lane 7 and 14, and **Figure 5-2 D** lane 5 and 13). A substantial amount of Prp22 can be bound under native conditions in the subsequent FLAG capture (compare lane 6/14, 7/15 with 8/16 in **Figure 5-2 D** for left/right panel). In the final *in vitro* spliceosome iCLIP preparation prior denaturation with 6 M urea was implemented as before. Although the amount of rPrp22-G512A added to the splicing extract far exceeds the WT levels, much more is lost during the whole purification procedure (**Figure 5-2 D** compare left and right panel). Despite this excess, even lower levels are eluted in a similar amount of spliceosomes (compare **Figure 5-2 C** lane 7/14 and **Figure 5-2 D**, lane 6/14), implying a decreased RNA binding ability of the dominant negative protein. The input concentration of rPrp22-K512A was adjusted to reach the same level as WT protein in purified spliceosomes. The final concentration in the extract thus reached a similar level as rPrp16-G378A for the complex C preparation (compare **chapter 3**).

5.1.2 Crosslinking and cDNA preparation suggests a reduced RNA binding ability for the dominant negative mutant of Prp22

After successful purification of spliceosomal complex C* and FLAG capture of Prp22, all further steps of the *in vitro* spliceosome iCLIP procedure can be performed according to the iCLIP protocol developed in **chapter 3**. Firstly, crosslinking of Prp22 to RNA had to be tested, and the RNase I concentration had to be optimised to produce suitable fragments for NGS. A dose of 312 x 100 µJ/cm^2 UV light is high enough to crosslink most protein-RNA complexes without over crosslinking the complex, which is indicated through a saturated radioactive signal (**Figure 5-3**). Comparing the different amounts of RNase I, digestion with 5 units produces the strongest signal fading from the protein size towards higher molecular weight species. A lower concentration would likely produce longer RNA fragments; however, the higher concentration is required to release an additional protein which crosslinks to the same RNA target (**Figure 5-3**, compare highlighted, lower molecular weight band). A larger amount of dominant negative rPrp22-K512A crosslinks to a lower amount of RNA than in the case of the WT protein, indicating that the mutant protein binds RNA less tightly (**Figure 5-3** compare radioactive blot (top) to western blot (bottom)). This observation

is unexpected as the ATPase deficient Prp16-G378A binds RNA more tightly than the WT protein (compare **Figure 3-7**), while the amino acid substitutions map to helicase motif I in both Prp16 and Prp22.

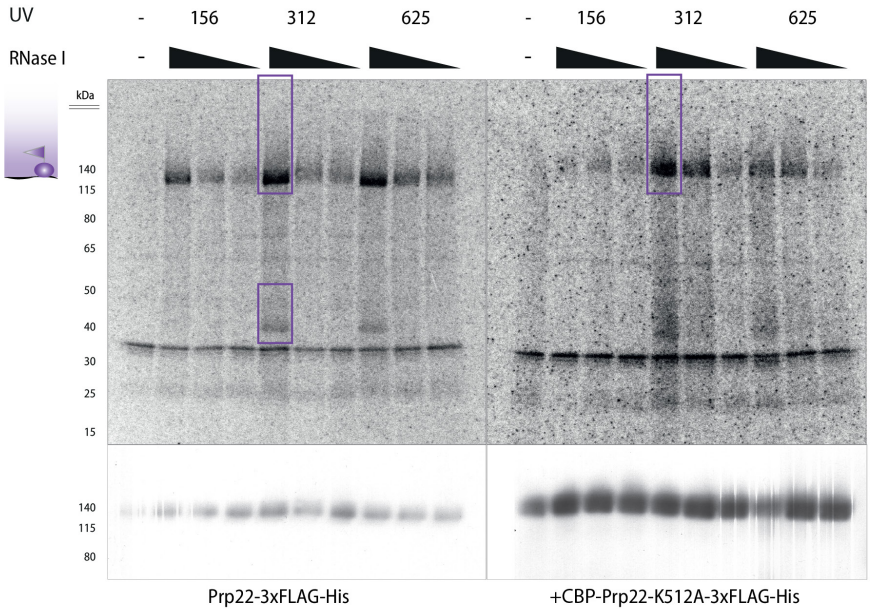


Figure 5-3: Prp22-WT binds RNA more tightly than rPrp22-K512A. Parallel fine-tuning of UV irradiation and RNase I digestion. Purified spliceosomal complex C* was irradiated with 156, 312, and 625 x 100 µJ/cm^2 UV light, followed by denaturation. The RNA was fragmented with 5, 0.5, and 0.05 units RNase I during FLAG capture. Subsequent γ-^{33}P-ATP radiolabelling allowed visualisation of bound RNA after SDS-PAGE separation and transfer onto nitrocellulose (top). The data presented derived from spliceosomes containing endogenous Prp22-3xFLAG-His (left) or rPrp22-3xFLAG-His (right), the highlighted lanes show the optimal condition which produces RNA-protein species from ~130 kDa upwards. The box at a lower molecular weight (~40-50 kDa) marks an additional crosslinked protein-RNA complex which is released at higher RNase concentrations. The left and right part of the image are derived from separate blots and were subsequently probed with anti-FLAG antibody to visualise and compare the amounts of Prp22 and rPrp22-K512A (bottom).

The optimised UV and RNase conditions were used to prepare cDNA libraries for both Prp22-WT and rPrp22-K512A in complex C*. cDNA amplification resulted repeatedly in short fragments on the TBE gel (data not shown). Regarding the sequencing data, hardly any reads aligned to the substrateome or the yeast genome (**Appendix folder**, session_11).

Prp22 proposed binding site included nucleotide 17 of the second exon in the *ACT1* transcript, which was determined by 4-thiouridine labelled RNA crosslinking in the context of a dominant negative Prp22 mutant protein (Schwer 2008). Assuming Prp22 binds at a similar location on the second exon of *UBC4* (here 25 nt long) of the purified complex C* in this work, less than 8 nucleotides would extend from the occluded site towards the 3'-end of the transcript. This extremely short distance could be the reason for the short cDNA fragments which are almost impossible to separate from primer- and adapter artefacts during gel purification. Given that a minimum length of 11 nt is

required for unique mapping in the bioinformatic pipeline, these experiments must be repeated with RNA substrates having a more extended second exon.

5.2 *In vitro* spliceosome iCLIP setup for Prp22 in complex P

In spliceosomal complex P, the intron-lariat is excised and the exons are ligated, while both of the products are still bound by the complex. To allow exon ligation, the WT version of *UBC4*_short is used during the splicing reaction for the *in vitro* spliceosome iCLIP procedure. Supplying dominant negative Prp22 stalls splicing at complex P (**Figure 5-4 A**). The recombinant protein rPrp22-K512A-3xFLAG, which was produced to locate Prp22 in complex C*, was used here, too. The ATPase-deficient mutant is incapable of removing the ligated exons efficiently and therefore hinders the progression towards ILS.

Using the previously shown purification strategy involving capture via the RNA substrate leads to a mixture of spliceosomal complexes predominated by complex C/C*. (Wilkinson et al. 2017) optimised the strategy by including specific antisense-DNA directed RNase H digestion targeting an undocked second exon when using *UBC4*_short_3xMS2 as splicing substrate. Precatalytic spliceosomes containing pre-mRNA and spliceosomes of the first catalytic step are ablated by this digestion, which is indicated by the absence of the pre-mRNA after incubation with the antisense oligonucleotide (**Figure 5-4 B**, compare lane 2 and 3). After amylose elution (lane 7), a mix of spliceosomal complex P, fragments of the RNase H digestion and possibly some free mRNA, released by residual activity of Prp22-K512A, are eluted. In *in vitro* spliceosome iCLIP, subsequent FLAG capture on Prp22-K512A ensures all data derives solely from complex P.

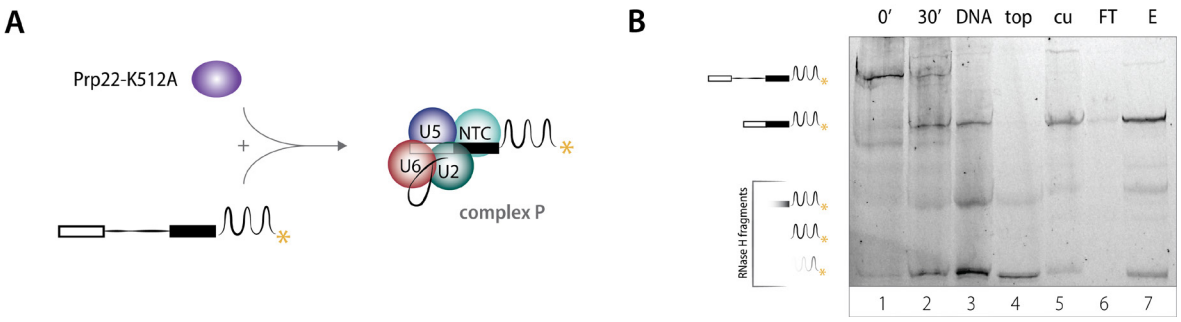


Figure 5-4: Purification of spliceosomal complex P. (**A**) The dominant negative rPrp22-K512A stalls the spliceosome at the complex P stage when assembled on WT pre-mRNA substrate, the yellow * indicates a fluorescein label. (**B**) RNA species are visualised by the 3'-fluorescein-label on a 10 % denaturing PAGE during complex P preparation. After 30 min incubation mRNA accumulates (0' and 30'). Incubation with the antisense oligonucleotide for endogenous RNase H digestion leads to the reduction of pre-mRNA species (lane 3). Spliceosomes bound to mRNA are rather found in the cushion (lane 5) than the top (lane 4) fraction after centrifugation through the glycerol cushion. The eluate (lane 7) of the amylose capture contains complexes bound to mRNA and fragments of the RNase H digest, whilst the flow-through (lane 6) is very clean. Purification of spliceosomal complex P essentially followed the protocol in (Wilkinson et al. 2017) with the exception of stalling with the dominant negative mutant rPrp22-K512A instead of rPrp22- S635A.

The optimal UV dose and RNase I amount was titrated to 312 x 100 µJ/cm^2 and 0.5 units, respectively (**Figure 5-3** shows and explains the titration in the context of complex C*). In contrast to the observation in complex C*, higher RNase I concentrations did not release co-crosslinking proteins. To retain longer RNA fragments, the lower concentration was adopted.

cDNA libraries were produced according to the *in vitro* spliceosome iCLIP protocol described in **chapter 3**. Data from the short CLIP protocol resulted in a binding profile independent of UV irradiation, further highlighting the lack of stringency during the short procedure. Most truncations accumulate on the second exon between nucleotides 8 to 26 (**Appendix folder**, session_12; read coverage in **Figure 5-5**).

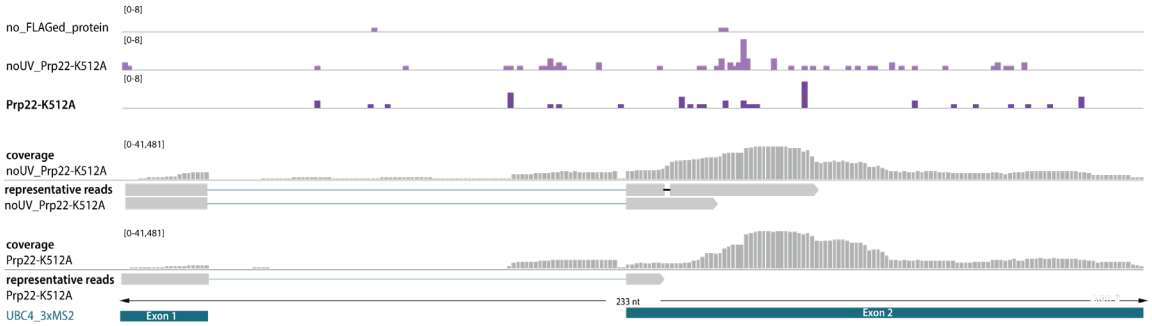


Figure 5-5: Mapping of Prp22 *in vitro* spliceosome iCLIP data to the substrate-ome for spliceosomal complex P.
First three lines: *in vitro* spliceosome iCLIP crosslinking data. Signal accumulates on the transcript. Both non-UV and irradiated samples accumulate truncations around nucleotide 12 and 41 of the second exon with a different distribution pattern. Further four lines: short-iCLIP experiments showing the coverage (reads/nucleobase) and some representative cDNA reads (grey horizontal bars with arrow head) for the indicated conditions that are derived from spliced RNA. The blue line between the grey bars implies that reads are spanning splice junctions. The black line marks a deletion in the mapped read. The coverage of short-iCLIP experiments is depicted here, as the coverage of the SDS-PAGE derived data (first three lines) was very low.

The method including separation by SDS-PAGE did not produce many uniquely mapped reads (**Figure 5-5**, first three lines, score 8; **Appendix folder** session_13). Since the used transcript is extended by ~100 nt through the MS2 stem loops at the 3-end here, read loss due to a length restriction can be excluded. Most reads map to the RNA transcript, with an accumulation of truncations between position 12 and 41 of the second exon (**Figure 5-5**, first three lines). Despite the different pattern derived from UV and non-UV irradiated samples, the difference in signal to background is too low to draw firm conclusions. Most aligned reads which span the length of the ligated exons, exclude the intronic sequence indicating homogenous sample preparation of complex P (**Figure 5-5**, last four lanes, short CLIP experiment). This experiment has to be optimised and repeated until a clean background signal will allow meaningful interpretation.

5.3 Towards time-resolved *in vitro* spliceosome iCLIP for Brr2 starting with complex B

5.3.1 Purification of spliceosomal complex B

Brr2 *in vitro* spliceosome iCLIP was initiated with the intention to follow helicase translocation during U4/U6 snRNA unwinding, starting from Brr2's pre-loaded position on U4 snRNA in complex B.

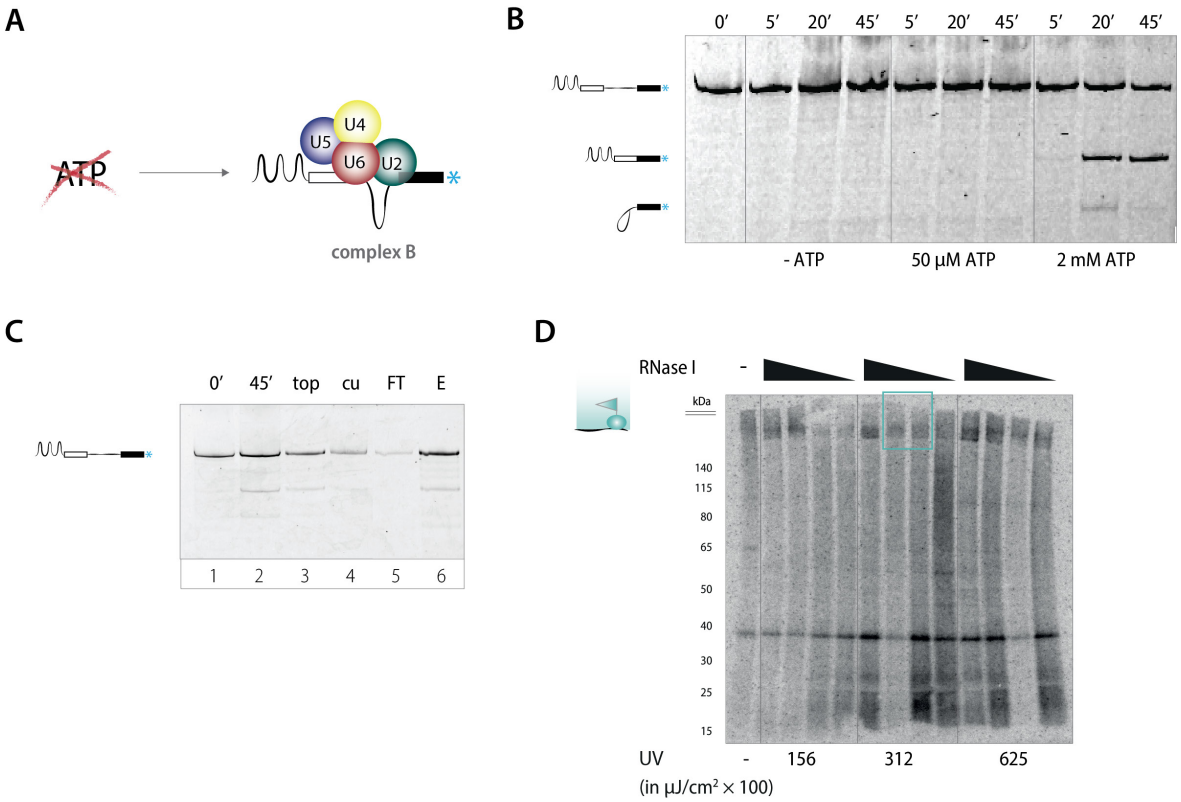


Figure 5-6: Purification and crosslinking of complex B. (**A**) Limiting ATP amounts stall spliceosome assembly at the complex B stage during *in vitro* splicing. For visualisation during the purification procedure, the pre-mRNA substrate is labelled with Cy5 (indicated with a blue *). (**B**) Time course for splicing reactions with no added ATP (left), 50 µM ATP (middle), and 2 mM ATP (right) separated on a 10 % denaturing PAGE. Splicing extract shows strong activity under normal splicing conditions with 2 mM ATP where about half the pre-mRNA is processed into mRNA after 20 min. Lowering the ATP concentration to 50 µM in the splicing reaction stalls the reaction as efficiently before the first catalytic step as performing splicing in the absence of ATP (compare 50 µM ATP and - ATP). (**C**) 10 % denaturing PAGE shows RNA species during complex B purification before the splicing reaction (0'), after the incubation under limiting ATP concentrations (45'), top (top) and bottom (cu) of the glycerol cushion, as well as flow-through and elution of the amylose capture. The band appearing after 45 min below pre-mRNA on the 10 % gel, does not correspond to mRNA or intron-lariat-intermediate, and only appears when the reaction is performed in an *in vitro* spliceosome iCLIP reaction scale with MS2-MBP. (**D**) Fine-adjusting of UV dose and RNase I. The splicing reaction was irradiated with 156, 312 or 625 x 100 µJ/cm^2 UV light as indicated. Glycerol gradient purification and amylose capture were performed before FLAG capture, when the RNA was fragmented by 50, 5, 0.5 or 0.05 units RNase I as indicated by the decreasing triangle. Radiolabelling with γ-^{33}P-ATP allowed visualisation of RNA after separating the complexes on SDS-PAGE and transfer onto nitrocellulose.

In a first step, a yeast strain was generated introducing the 3xFLAG-polyhistidine-tag at the C-terminus of endogenous Brr2. As described in the literature (Tarn, Lee, and Cheng 1993; Chan et al.

2003) limiting ATP concentration to 50 µM stalls a splicing reaction at the complex B conformation. This strategy was successfully used to purify spliceosomal complex B for structural analysis (Plaschka, Lin, and Nagai 2017), and will be employed to generate complex B for *in vitro* spliceosome iCLIP, too (**Figure 5-6 A**).

Splicing extract derived from Brr2-3xFLAG-His_8 cells was extensively dialysed to remove most ATP. I tested three additional techniques to remove residual ATP, which all showed the desired effect: (1) prior incubation of the extract with 2 mM glucose to deplete ATP as in (Abovich, Liao, and Rosbash 1994), (2) buffer exchange of the extract via G25 column removing small molecules including ATP, and (3) performing the splicing reaction additionally under limiting magnesium concentrations of 300 µM (Bertram et al. 2017) (data not shown). Finally, I implemented pre-incubation with glucose into my protocol for complex B preparation. Typical activity under usual splicing conditions was observed, whereas the reaction was arrested before the first catalytic step under limited ATP concentrations (**Figure 5-6 B**, compare 2 mM and 50 µM ATP). Subsequent purification via glycerol gradient and amylose capture produced a mostly homogeneous sample (**Figure 5-6 C**, lane 6).

5.3.2 Brr2 can be crosslinked to RNA under low ATP conditions in splicing extract

Brr2 is pre-loaded onto U4 snRNA in complex B (Plaschka, Lin, and Nagai 2017), and is then activated to unwind the U4/U6 snRNA duplex by an unknown signal, which in turn will result in a major rearrangement and exchange of proteins to produce spliceosomal complex B^{act}. To observe Brr2 translocation in the most controllable and also most natural format, spliceosomes shall be chased and UV-irradiated in extract for time-resolved *in vitro* spliceosome iCLIP. Thus, all required components for remodelling such as the extended NTR/NTC as well as unknown factors will be provided by the extract.

Therefore, after assembly of spliceosomal complex B at 50 µM ATP, the extract was irradiated to compare different UV doses. Following purification by glycerol gradient and amylose capture, the crosslinked complex was subjected to the iCLIP procedure including denaturation by urea, FLAG capture, and finding the right concentration of RNase I for fragmentation. Optimal conditions, in which RNA fragmentation seemed to give the desired pattern, are achieved by irradiation with 312 x 100 µJ/cm^2 UV light and a condition between 5 and 0.5 units RNase I (**Figure 5-6**). Additional experiments to optimise the RNase I concentration further did not give a clearer indication for optimal amounts.

Therefore, cDNA libraries were produced from both conditions to assess the data quality at the start point of Brr2 unwinding. The experimental set-up used is comparable to the one used later to follow remodelling but slightly shorter. Unfortunately, not many reads could be mapped at all from

this experiment to the substrate-ome and reads which locate to the yeast genome seem to terminate randomly scattered (**Appendix folder**, session_14).

5.3.3 Purified complex B can be chased to later stage spliceosomes in splicing extract

To provide controlled conditions for a chase the following was considered: complex B shall be purified from extract containing the 3xFLAG-tagged Brr2 and then be transferred to a different splicing extract which provides additional factors required for the progression of the splicing reaction while inhibiting the exchange of snRNPs and helicases.

Having a FLAG-tagged version of Brr2 only in the first extract ensures that iCLIP data is solely derived from Brr2 of the pre-assembled complex B. Treatment of splicing extract with micrococcal nuclease was shown to prevent splicing by degrading snRNAs (Cheng and Abelson 1986; Yeh et al. 2011) and can therefore be included to prevent snRNP exchange in the second splicing reaction.

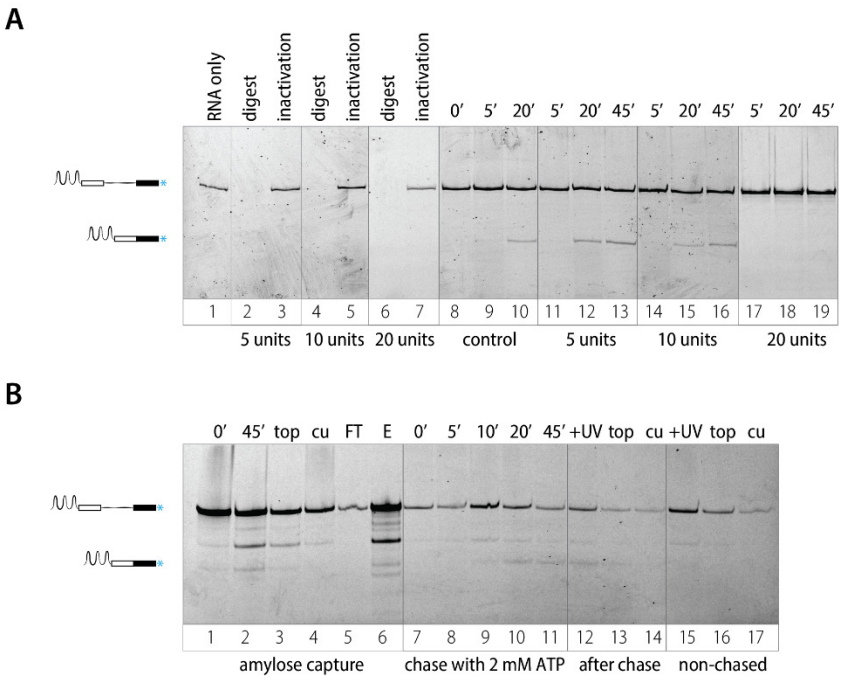


Figure 5-7: Complex B can be chased in micrococcal nuclease treated extract. (**A**) Splicing extract was treated with 5, 10, or 20 units nuclease per 1 µL splicing extract in the presence of 2 mM $CaCl_2$ at 30 °C. Lanes 2, 4 and 6 (digest) contained RNA before the enzymatic reaction. For lanes 3, 5, and 7 (inactivation), the extract was treated with the nuclease, inactivated with 8 mM EGTA and then incubated with RNA to test for residual nuclease activity. Likewise, treated extract was subjected to a time course of a splicing reaction to screen for efficient removal of splicing activity (lanes 8-19). Lanes 8-10 show a control time course without nuclease treatment. Parts of this image are derived from separate gels. (**B**) Spliceosomal complex B was assembled under 50 µM ATP for 45 min, applied to a glycerol cushion, followed by amylose capture. The amylose eluate was subjected to a further splicing reaction containing either 50 µM (no timepoints taken) or 2 mM ATP (lanes 7-11). The splicing extract of the second reaction was pre-treated with micrococcal nuclease. Both conditions were irradiated with 312 x 100 µJ/cm^2 UV light (lane 12 for 2 mM ATP; lane 15 for 50 µM ATP) and applied to another glycerol gradient (lanes 13/14 and 16/17). The denaturing 10 % PAGE shows the RNA-species by Cy5-label.

Micrococcal nuclease requires Ca^{2+} for its activity and can be inactivated by the chelating agent ethylene glycol-bis(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid (EGTA). Complete inhibition of splicing activity while preventing digestion of added RNA after EGTA inactivation is a measure to find optimal conditions. Full digestion of the RNA amount used for an *in vitro* splicing reaction (1.5 fmol / 1 µL extract) was obtained with 5 units nuclease per 1 µL splicing extract supplemented with 2 mM Ca^{2+} (**Figure 5-7 A** lane 2). A concentration of 8 mM EGTA was required to block the activity of the nuclease completely (final concentration shown in **Figure 5-7 A** inactivation). A concentration of 20 units/(µL extract) is needed to block the splicing activity, meaning extract treated in this way will be unable to provide intact snRNPs (lanes 17-19).

In a second step, chasing of purified spliceosomal complex B was tested in the nuclease-treated extract. The first purification under limiting ATP concentrations produced spliceosomal complexes mainly bound to pre-mRNA as expected for complex B (**Figure 5-7 B**, lane 6). A second splicing reaction set up with the purified spliceosomes and micrococcal nuclease-treated extract shows the formation of mRNA after 10 min incubation under standard splicing conditions (2 mM ATP) (**Figure 5-7 B**, lane 9). After 45 min about a quarter of pre-mRNA is turned into mRNA (lane 11) indicating that complex B can be chased to later spliceosomal complexes under these conditions.

A control experiment was performed with limiting ATP concentrations (50 µM ATP) during the second splicing reaction. Assembled spliceosomes obtained from both conditions were crosslinked with the dose determined in the previous **subchapter 5.3.2**, and then applied to another glycerol cushion and denatured by 2 M urea before FLAG capture to separate Brr2-crosslinked species. After the iCLIP specific steps and visualization by radioactive RNA labelling, both conditions showed strong background lacking the specific signal for Brr2-RNA complexes despite giving a clear signal for Brr2 protein in western blot (data not shown).

The high background can be explained as crude extract is loaded onto FLAG beads and denatured with a lower concentration of urea compared to the previously described *in vitro* spliceosome iCLIP method (**chapter 3**). Increasing urea to 6 M followed by dialysis to allow FLAG capture was no improvement. In a different format, complex B was retained on the amylose beads during the second splicing reaction aiming to get a cleaner result by applying more purified components onto the FLAG beads. However, the rearrangement of complex B to later spliceosomal complexes was massively hindered on beads: hardly any mRNA was produced during the chase (data not shown).

Successful generation of crosslinked Brr2-RNA species by irradiation of a splicing reaction implies crosslinking should have been successful in these conditions, too and the signal might simply be obscured by non-specifically captured RNA (compare previous **chapter 5.3.2** and **Figure 5-6 D**). For

future experiments, the MS2 binding protein will be coupled to an additional tag to allow two subsequent steps of capture on the RNA substrate, thereby avoiding the need to load crude extract onto FLAG beads.

In summary, chasing spliceosomal complex B towards the catalytic stages in inactivated extract while crosslinking Brr2 seems feasible. Finding conditions providing a suitable time scale to allow time resolved crosslinking, however, will be another challenging task.

5.4 Chasing Prp16 in C complex

5.4.1 Prp16 can be depleted from splicing extract by its epitope tag

Previously, Prp16-mediated transition of spliceosomes at the complex C state was observed by FRET (Semlow et al. 2016). In the mentioned study, ΔPrp16 spliceosomes were incubated with recombinant Prp16, ATP and magnesium while the conformational change of the RNA transcript was observed by FRET. This clearly indicates that a rearrangement of the RNA network can be obtained by the action of Prp16 itself. Here, we want to investigate this Prp16-mediated transition by visualising Prp16's path on the RNA via *in vitro* spliceosome iCLIP. Next, a complete transition spanning the rearrangement towards the second step might be investigated by adding second step factors such as Slu7, Prp18, and Prp22 from recombinant source.

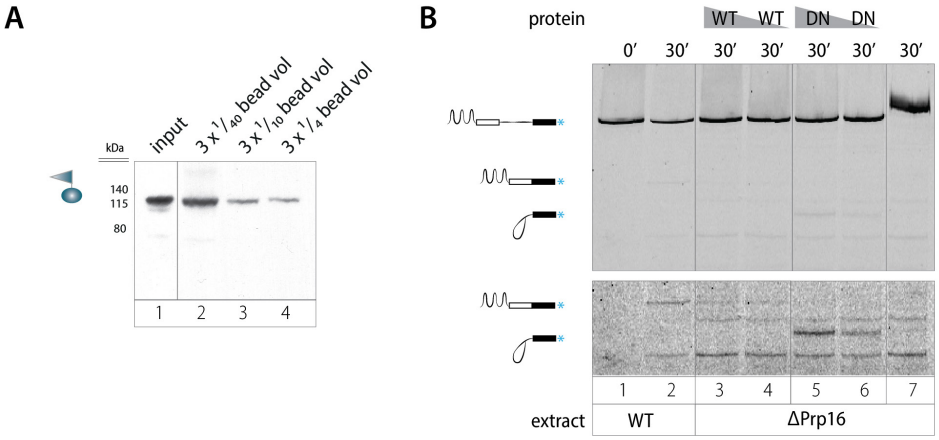


Figure 5-8: Prp16 depletion from splicing extract. (**A**) Protein amounts are compared between the original extract (10-fold dilution) and extracts depleted of Prp16-3xFLAG by FLAG capture. Incubation of the extract thrice with either 1/40th, 1/10th or 1/4th volume of beads strongly reduces the protein amount. (**B**) Splicing test for WT extract (expressing Prp16-3xFLAG, non-depleted, lanes 1 and 2) and depleted from Prp16 (ΔPrp16, lanes 3-7, extract derived from lane 4 in (**A**)). Addition of two concentrations rPrp16-G378A (lanes 5,6), addition of two concentrations rPrp16-WT (lanes 3, 4), no addition of recombinant protein (lane 7). The overall splicing activity of the extract is very low. The bottom panel shows only the lower part of the gel with increased contrast.

To purify ΔPrp16 spliceosomes, Prp16 has to be depleted from splicing extract. In contrast to using antibodies raised against Prp16, here the extract is depleted using the endogenously introduced

C-terminal epitope tag and magnetic beads coupled to the complementary antibody. In an initial trial, Prp16 was depleted from yeast extract expressing endogenous Prp16-3xFLAG. To rescue the loss of splicing activity of the depleted extract, rPrp16-WT-3xFLAG was expressed and purified (**7.3.1**). Three consecutive incubations of splicing extract with 1/4th bead volume reached about 100-fold reduction of the protein (**Figure 5-8 A**, compare lane 4 to lane 1). When comparing this extract to the original extract, it does not build up mRNA implying a second step splicing defect (**Figure 5-8 B**, compare lane 7 to lane 2). When supplementing the splicing extract with rPrp16-G378A, this phenotype is enhanced, while mRNA is produced again when supplemented with rPrp16-WT (**Figure 5-8 B**, lanes 5/6, 3/4). The levels of mRNA obtained from WT extract, however, cannot be reached indicating insufficient amounts of rPrp16 were added back to the extract or the antibody depletion of the Prp16 influences the second step of splicing in an additional way by e.g. co-depleting other essential factors. The depletion experiment has to be improved to rule out this possibility. For *in vitro* spliceosome iCLIP, Prp16 would ideally have an epitope tag for depletion different to the one on the rPrp16-WT which is used to chase the complex and captured during iCLIP immunoprecipitation. To fulfil this requirement, a yeast strain expressing Prp16-3xHA is currently under construction.

5.4.2 Prp16 binding to RNA is ATP-independent and compatible with pre-steady state ensemble kinetics

Following the path of a helicase with *in vitro* spliceosome iCLIP would be a bulk kinetic experiment, meaning several molecules are observed together while the translocation of a single helicase molecule will be averaged out by others. In contrast, single molecule experiments can distinguish between each molecule and are therefore more accurate.

In so called pre-steady state ensemble kinetic experiments, differences in translocation velocity are kept to a minimum. The molecules are synchronised at the start of the experiment, which provides a short time window to observe translocating helicases which are still synchronised (Fischer and Lohman 2004). Rebinding of helicases to the target RNA would lead to incorrect assumptions about kinetics and processivity. In single-turnover experiments a protein trap such as heparin, poly-U RNA or the natural substrate will catch unbound helicases before rebinding the target RNA (Lucius et al. 2003). This subchapter aims at determining pre-steady state and single-turnover conditions with the intention of following the path of Prp16 in ΔPrp16 spliceosomes.

Initially, binding properties of Prp16 should determine if the helicase needs ATP and/or magnesium for RNA binding in order to find a pre-steady state condition in which rPrp16 is bound to ΔPrp16 spliceosomes but unable to translocate. With most techniques, binding of the comparatively small

protein Prp16 to the spliceosome cannot be monitored. Therefore, the binding of rPrp16 to single stranded RNA was determined by polarisation experiments.

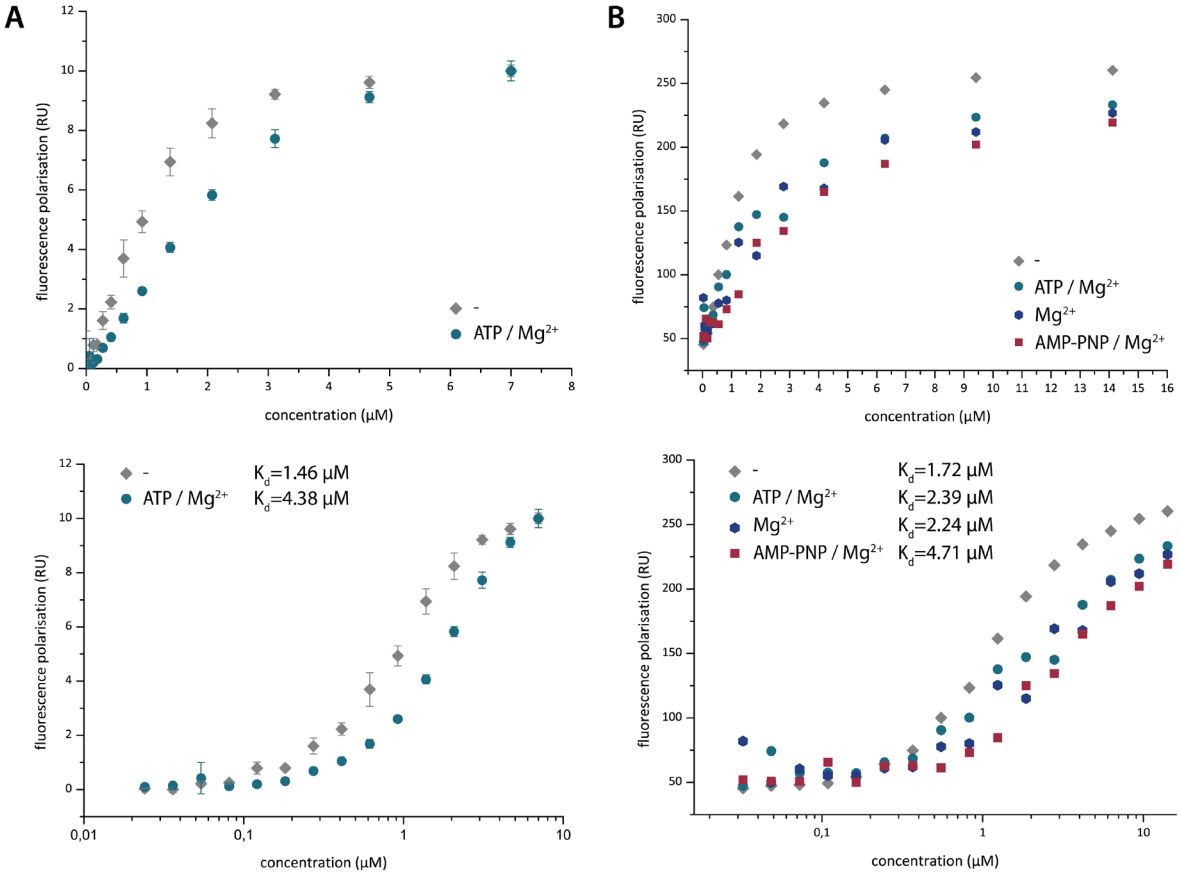


Figure 5-9: Fluorescence polarisation determines low micromolar K_d for binding of Prp16 to poly-U RNA. (**A**) Dilution series of Prp16 to bind fluorescein-labelled poly-U RNA in reaction buffer with and without the addition of ATP and Mg^{2+} (triplicates, error bars show standard deviation). (**B**) Dilution series of Prp16 to bind fluorescein-labelled poly-U RNA under the following conditions: no addition, +ATP/Mg^{2+}, +Mg^{2+}, +AMP-PNP/Mg^{2+}. (**A**) and (**B**) binding buffer contains 100 mM KCl. The data points were fitted to a one-site specific binding event with the following formula Y=Bmax*x/(K_d + x) in GraphPad Prism. Bmax = maximum specific binding (same unit as Y-axes). The obtained K_d-values are indicated in the graph. Top: linear representation of the data. Bottom: logarithmic representation.

The K_d for binding fluorescein-labelled poly-U RNA was calculated to the low µM-range by fitting the polarisation values, measured over an increasing concentration of rPrp16, to a one-site specific binding event curve (**Figure 5-9**). Complete saturation could not be reached, since rPrp16 precipitated at concentrations higher than ~30 µM. Comparing the influence of ATP and Mg^{2+}, the K_d was lowest in the absence of both (**Figure 5-9**). The addition of ATP in combination with Mg^{2+} reduces the binding affinity slightly, which could indicate that the helicase starts translocating or loses contact with the RNA substrate. The same effect can be observed by addition of solely Mg^{2+} or the non-hydrolysable ATP analogue AMP-PNP in combination with Mg^{2+} (**Figure 5-9 B**). The reduced binding affinity through Mg^{2+}, ATP/Mg^{2+} or AMP-PNP/Mg^{2+}, might indicate the recombinant helicase is

bound to ATP already. Taken together, these results indicate that rPrp16 can bind RNA without the external addition of ATP or Mg^{2+} which in turn means the helicase can be bound to the stalled ΔPrp16 spliceosome and the reaction can be started with the addition of either ATP/Mg^{2+} or Mg^{2+} in a pre-steady state ensemble kinetic format. In addition, the experiment clearly shows that poly-U RNA can be used as a trap to ensure single-turnover.

A more accurate K_d could be determined through switchSENSE technologies, in which association rate (k_{on}) and dissociation rate (k_{off}) are measured separately. The measurements would allow one to judge whether a decrease in affinity derives from an increased off-rate and vice-versa. Since setting up these experiments only provides secondary information but require a lot of optimisation, this option will be considered when needed.

5.5 *In vitro* spliceosomal iCLIP can be transferred to additional helicases

In summary, this chapter shows that the DEAH-box helicase Prp22 can be analysed by *in vitro* spliceosome iCLIP in a similar way to the closely related Prp16. Hardly any cDNAs were generated and uniquely mapped to the target transcript in the context of spliceosomal complex C*, possibly because Prp22 might locate near the 3'-end of the artificial pre-mRNA substrate. Assembling the spliceosome on *UBC4* transcript with the extended second exon (61 nt) or the *ACT1* transcript, which both improved mapping for Prp16 *in vitro* iCLIP data (compare **chapter 4**), is expected to generate longer reads and might therefore result in a clear binding profile for Prp22.

In a similar way, the DEAH-box helicases Prp2 and Prp43 could be analysed by *in vitro* spliceosome iCLIP as soon as purification methods for spliceosomal complexes bound by the respective helicase are established. Prp43 is reported to be involved in disassembly of different stage spliceosomes arrested when remodelling by Prp16 or Prp22 is hindered (Mayas, Maita, and Staley 2006; Koodathingal et al. 2010). Obtaining the binding profile of all helicases, would allow for comparison of binding properties within and between different families of helicases as well as between different spliceosomal complexes.

Furthermore, this chapter gives first indications for possible time resolved *in vitro* spliceosome iCLIP formats and their potential workflow is summarised in **Figure 5-10**.

For both Brr2 and Prp16 time-resolved *in vitro* spliceosome iCLIP, finding the right time-scale will be the most challenging task. It might require performing the chase with slow- or non-hydrolysable ATP, low magnesium concentrations or low temperatures. The optimal time frame can most likely only be determined by the comparison of final binding profiles. Of course, it remains unclear if unwinding events can even be captured by UV irradiation. The process might instead lead to a broader peak distribution or the short-lived events will only allow capture of the start and/or end points.

This chapter has certainly shown, that there is a potential for time-resolved *in vitro* spliceosome iCLIP, however a lot of work still has to be done. In the future, combining the binding profile of different helicases might enable one to follow the entire spliceosome remodelling process from assembly to activation, catalysis and disassembly from the viewpoint of the spliceosomal helicases.

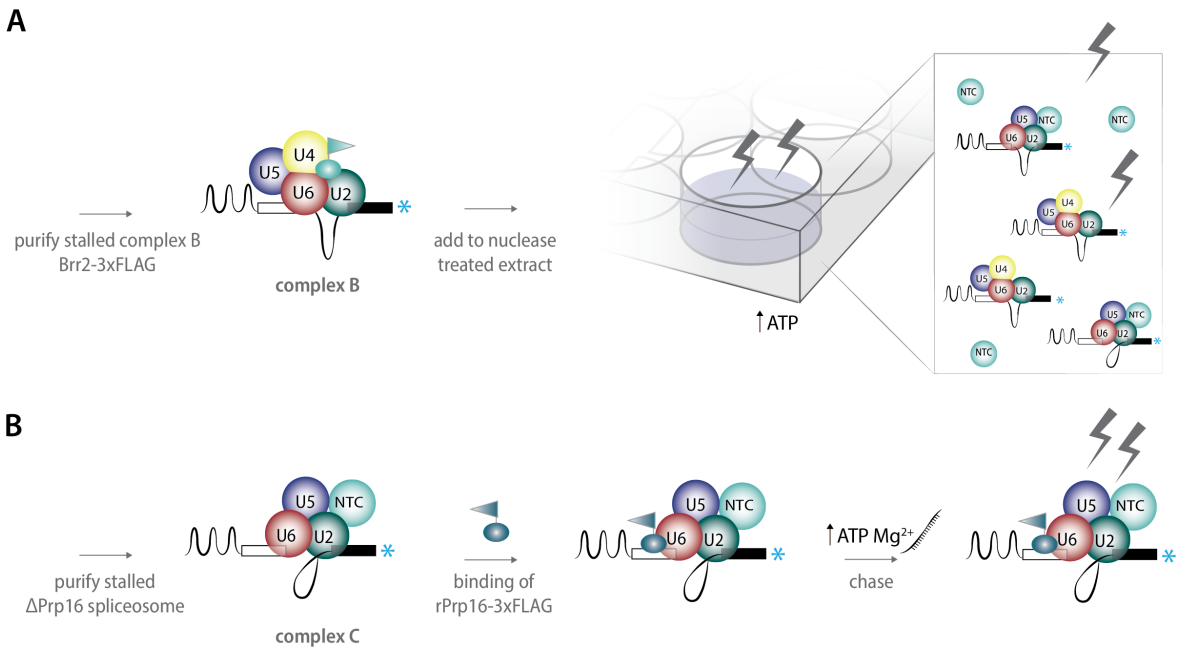


Figure 5-10: Possible formats for time-resolved Brr2 and Prp16 *in vitro* spliceosome iCLIP. (**A**) Spliceosomal complex B is purified under limited ATP conditions from Brr2-3xFLAG splicing extract. The complex is transferred to micrococcal nuclease treated extract and splicing is initiated by increasing ATP. While the assembled complex B remodels into later spliceosomal complexes, the extract is irradiated with UV light at different time points. The previously described iCLIP procedure will then identify the path of Brr2 while unwinding the U4/U6 snRNA duplex. (**B**) Spliceosomal complex C is purified from Prp16-depleted extract. After binding of rPrp16-3xFLAG to the complex, the chase can be initiated by ATP and magnesium. Addition of a protein trap such a poly-U RNA will prevent rebinding of the same molecule to the spliceosome. During the chase UV irradiation over time will capture the translocation of Prp16 which can be visualised in the binding profile obtained by the iCLIP procedure.

6 Discussion

The spliceosome plays a major role during eukaryotic gene expression by determining the amount and composition of expressed proteins and non-coding RNAs (Pawlicki and Steitz 2010). Understanding the mechanism of spliceosome-catalysed pre-mRNA splicing in detail is crucial to explain post-transcriptional gene regulation through transcript maturation in the nucleus.

In recent years, our understanding of the spliceosome benefitted greatly from snapshots taken during the splicing process picturing both composition and structure of the spliceosome (Fica and Nagai 2017) as well as profiling of spliceosome substrates in a genome-wide manner (Burke et al. 2018; Chen et al. 2018). Despite being objectives of many biochemical studies, it is crucial to study the extensive remodelling events, responsible for conversion between different spliceosomal states, before spliceosome catalysed splicing is thoroughly understood (Makarov et al. 2002; Schwer 2008; Mefford and Staley 2009; Hoskins, Gelles, and Moore 2011; Tseng, Liu, and Cheng 2011; Krishnan et al. 2013; Semlow et al. 2016; Hahn et al. 2012; Boesler et al. 2016). As described in the introduction, the major initiators of remodelling are the spliceosomal helicases, which in cryoEM structures are often poorly resolved due to their peripheral location, making mechanistic conclusions difficult even in a static spliceosomal particle. During the course of this thesis, *in vitro* spliceosome iCLIP was developed as a tool to investigate spliceosomal helicases biochemically within the spliceosome.

The following chapter will concentrate on discussing and giving future perspectives on *in vitro* spliceosome iCLIP (compare **chapters 3**, **4**, and **5**). Concluding remarks for the chapter dealing with the spliceosomal protein Snu114 can be found at the end of the respective chapter (**2.3**).

6.1 *In vitro* spliceosome iCLIP is a versatile tool to study spliceosome remodelling

In vitro spliceosome iCLIP applies the NGS technique iCLIP to fully assembled spliceosomes to obtain the complete RNA-binding profile of even transiently bound spliceosomal helicases within the spliceosome.

A full pipeline for *in vitro* spliceosome iCLIP was initially developed to target two proteins in spliceosomal complex C: I have chosen Prp16 and SmB as the first target helicase and control protein to obtain binding profiles, respectively. SmB binding sites correlate with the known Sm sites and thereby validated the method (**Figure 3-9** and (Guthrie and Patterson 1988)). The final workflow for *in vitro* spliceosome iCLIP is summarised in **Figure 6-1**.

The starting material for the method consists of specifically stalled and purified spliceosomes, which were assembled from yeast extract onto defined pre-mRNA substrates. The experimental setup allows one to draw mechanistic conclusions through variation of the pre-mRNA substrate and the comparison of binding profiles between helicase mutants.

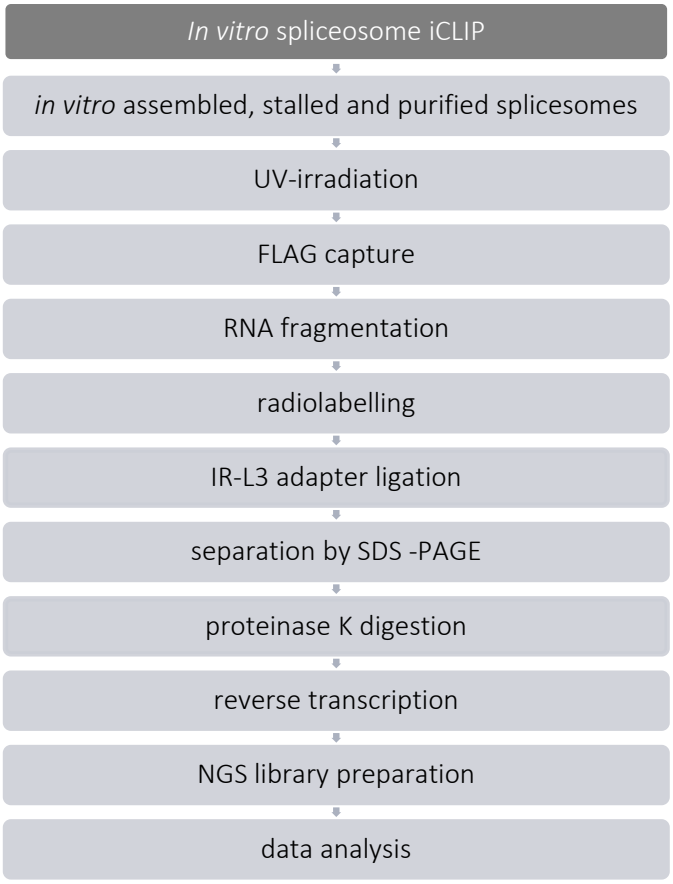


Figure 6-1: Workflow of *in vitro* spliceosome iCLIP. *In vitro* assembled and stalled spliceosomes containing a FLAG-tagged target protein are purified and irradiated with 254 nm UV light to induce covalent crosslinking between RNA and protein. The UV amount has to be adjusted anew for every target protein. Sometimes denaturation of the complex is necessary to liberate the FLAG-tag. During FLAG-capture, controlled RNase I digestion fragments the RNA suitable for NGS; another step which requires adjustment for every target protein. Since RNase I seemed to be best for fragmentation, 3'-ends have to be dephosphorylated. The 5'-end is radiolabelled to support visualisation of the RNA-protein complex and the 3'-end is linked to the IR-L3 adapter allowing later visualisation and RT primer annealing. The protein-RNA complexes are separated by SDS-PAGE, blotted onto nitrocellulose, and extracted by proteinase K digestion. Reverse transcription truncates when the enzyme reaches the crosslink site which is occupied by the remaining peptide. Amplification and purification steps produce a cDNA library for single read sequencing with the Illumina HiSeq 2500 system. Data analysis identifies the crosslink sites and their abundance on the substrate-ome (splicing substrate and spliceosomal snRNAs) and the yeast genome.

Besides its variable setup concerning pre-mRNA substrate and helicase mutants, *in vitro* spliceosome iCLIP has the following advantages: Only direct protein-RNA contacts are observed due to the required proximity for efficient UV-crosslinking. By using the intrinsic crosslinking ability of RNA, base modifications in the substrate RNA or in endogenous snRNAs are avoided. A full binding profile is obtained from one experiment, in contrast to e.g. traditional crosslinking with photoreactive derivatives *in vitro*, in which normally only a small number of specific nucleotides are tested one after the other for interaction. Binding to all RNAs, contained in the spliceosome at the given spliceosomal stage, is captured simultaneously and the RNAs are unambiguously identified by sequencing. The method can be further developed to be combined with other sequencing approaches such as hiCLIP, which identifies not only the target RNA region but also determines RNA secondary structure in the vicinity (Sugimoto et al. 2015). Finally, *in vitro* spliceosome iCLIP can be adapted to other spliceosomal helicases which act at various stages during splicing as shown through preliminary experiments in **chapter 5**.

Like every method, *in vitro* spliceosome iCLIP has several restrictions: Uridines crosslink with higher efficiency to proteins than other nucleotides introducing bias (Sugimoto et al. 2015). Secondary structure elements might crosslink with different efficiency than single-stranded unstructured RNA regions. The overall experimental procedure consists of many time-consuming steps, which are not automated and therefore require extensive hands-on time. In addition, we are still working on reducing the background signal from non-UV irradiated samples to increase the confidence in the identification of genuine crosslinks at nucleotide resolution.

By generating a binding profile of a target protein on the RNA, *in vitro* spliceosome iCLIP cannot reveal information on which parts of the protein interact with the RNA, nor where the protein is located in the spliceosome.

6.1.1 Specific and general obstacles for *in vitro* spliceosome iCLIP

While developing and analysing the binding profile for Prp16 in complex C, several issues arose, while most could be resolved (compare **chapter 4**). Although the obstacles were found in this specific setup, some of them are generally worth considering before setting up *in vitro* spliceosome iCLIP.

(1) *In vitro* spliceosome iCLIP is restricted by the number of nucleotides downstream of the crosslink site. In the binding profile of Prp16, the major crosslink site was found around 40 nt from the 3 -end of the splicing substrate used, which directly restricts the generation of uniquely mapped reads. A ~2-fold increase in the distance between the location of the major crosslink site and the 3'-end of the RNA substrate led to a 14-fold increase in the histo-

gram height (**Figure 4-4**; histogram height = number of reads with the same crosslink position). To avoid read loss, one should ideally aim to use an RNA substrate with an adequate extension at the 3'-end to allow mapping of the maximal 90 nt long reads (for read length in the included experiments, compare **Appendix Figure 8-3**).

(2) *In vitro* spliceosome iCLIP on spliceosomes of the first catalytic step is prone to false positive truncation events at the intron lariat-junction. Forming the unusual 2'-5' phosphodiester bond, both the brA and the 5'SS location become obstacles for the reverse transcriptase. In the binding profile of Prp16, both positions showed sharp peaks (**Figure 4-3** and **Figure 4-4**). We could show that nucleobases at both positions are, in contrast to genuine crosslink positions, often mutated after reverse transcription (**Figure 4-5 A**). In addition, the sharp peak at the brA could be decreased and the coverage upstream the brA increased by including an enzymatic debranching step in the *in vitro* spliceosome iCLIP procedure (**Figure 4-7** and **Figure 4-8**). Consequently, one should consider RNA-induced termination events, when performing sequencing approaches on the spliceosome.

(3) *In vitro* spliceosome iCLIP produces high background signal. Control samples, which derived from non-UV irradiated samples, resemble the binding profile of Prp16 in lower intensity (**Figure 4-3** and **Figure 4-4**). We could show that the region harbouring most crosslinks was clearly UV-dependent (**Figure 4-5**). Truncation events may be derived from very strong binding of the target protein to RNA, which is supported by signal on SDS-PAGE in the absence of UV irradiation (**Figure 3-7**). Possibly, the average 30 nt-long reads which are generated during *in vitro* spliceosome iCLIP accumulate around the binding site due to the strong RNA-protein interaction and then resemble crosslink events in the same region. Aiming to retain longer reads and applying even more stringent washing conditions will hopefully decrease the background level.

6.1.2 *In vitro* spliceosome iCLIP supports the model for Prp16 acting from a distance

In vitro spliceosome iCLIP data revealed the binding site for Prp16 in complex C. Crosslinks were found only on the splicing substrate, not on any of the spliceosomal snRNAs (**Figure 3-10**). The same binding profile was produced from the following three conditions: (1) Prp16-WT on AC (3'SS) substrate and the ATPase-deficient helicase Prp16-G378A on both (2) unmodified and (3) AC (3'SS) splicing substrate (**Figure 4-4** and **Figure 4-9**). Since the mutant protein has a residual ATPase activity of only 2 % (Schneider, Hotz, and Schwer 2002) which should severely restrict its helicase activity, the crosslink profile should reflect the starting point before Prp16 starts translocating to remodel spliceosomal complex C. To draw sequence and substrate-independent conclusions, the experiments were performed with both *UBC4* and *ACT1* pre-mRNA substrate. The major interaction site

for Prp16 was found to be about 15 nucleotides downstream of the brA independent of the substrate (**Figure 4-11**). Our data is in good agreement with previously determined positions from 4-thiouridine photo crosslinking (McPheeters and Muhlenkamp 2003; Semlow et al. 2016). Prp16's binding site seems to follow a molecular ruler mechanism, in which the binding distance from the brA is determined by the dimensions of the spliceosome body. In the cryoEM structure of complex C, the distance between the brA and Prp16 binding site was reported to span approximately 18 nt (Galej et al. 2016). Consequently, our data supports the model of Prp16 acting from its peripheral location to remodel distant RNA structures located in the catalytic core of the spliceosome. The model suggests that the translocation movement is transmitted into a pulling action when the helicase is physically obstructed by spliceosomal components (Semlow et al. 2016). Performing *in vitro* spliceosome iCLIP on *UBC4* or *ACT1* splicing substrates with artificial extensions between the brA and Prp16's binding site, could be used to validate the distance.

6.1.3 Pilot experiments on Brr2 and Prp16 suggest two strategies towards time-resolved *in vitro* spliceosome iCLIP

In addition to observing static interactions of spliceosomal helicases with their target RNA, we had the aspiration to modify *in vitro* spliceosome iCLIP towards a time-resolved method. Only binding profiles recorded over the course of the remodelling event have the potential to reveal the range of the helicases' translocation. Pilot experiments for the two helicases Brr2 and Prp16 suggest applying different strategies (**Figure 6-2**).

(1) In spliceosomal complex B, Brr2 is pre-loaded onto U4 snRNA (Plaschka, Lin, and Nagai 2017). To monitor Brr2 helicase translocation during the transition to complex B^{act}, purified complex B will be incubated in splicing extract, which provides all the factors required. While the assembled complex B remodels into later spliceosomal complexes, the extract is irradiated with UV light (**Figure 6-2** left).

We could already show that RNA-Brr2 crosslinks can be generated in stalled spliceosomal complex B when splicing extract was irradiated with UV light (**Figure 5-6 D**). In addition, we could show that purified complex B can remodel into later stage spliceosomes in micrococcal nuclease inactivated extract (**Figure 5-7 B**). The nuclease treatment is necessary to prohibit interchange of snRNPs between preassembled complexes and the splicing extract.

(2) To follow Prp16's translocation on RNA, we aimed to develop a system starting from Prp16 depleted spliceosomes, similar to the method described previously in (Semlow et al. 2016). We could show that depletion of Prp16-3xFLAG by epitope-antibody interactions, almost completely inhibits the second splicing reaction, stalling at the complex C stage (**Figure 5-8**).

In fluorescence polarisation experiments, we determined similar affinities for rPrp16-WT binding to poly-U RNA in the presence and absence of ATP/Mg^{2+} (**Figure 5-9**). This finding is crucial to establish pre-steady state conditions, in which Prp16 will firstly be bound onto the spliceosome and translocation initiated by the addition of the cofactors (Lucius et al. 2003). Poly-U RNA can then be used as a protein trap catching released helicases to guarantee single turnover conditions.

Once the pre-steady state set-up is developed to monitor Prp16 translocation, one could potentially determine Prp16 binding profiles for a complete transition to post-catalytic spliceosomes by providing second step factors such as Slu7, Prp18, and Prp22 from recombinant source (**Figure 6-2** right).

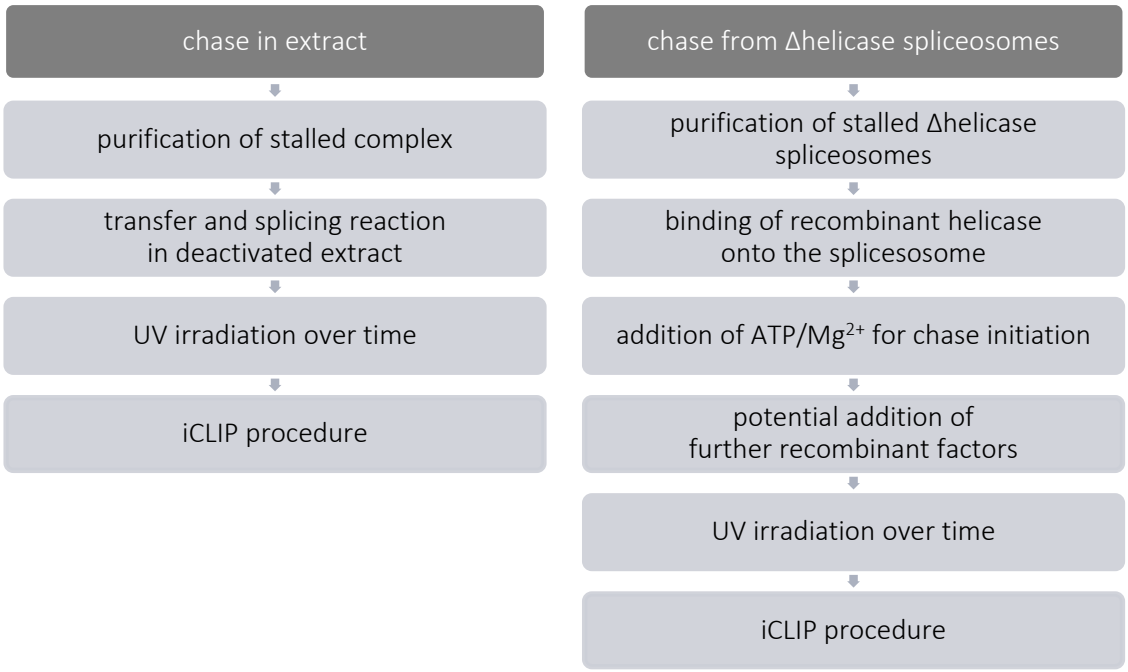


Figure 6-2: Potential workflow for time-resolved *in vitro* spliceosome iCLIP. (**left**) *In vitro* spliceosome iCLIP for time resolved profiles during a transition which is experimentally best performed in extract (e.g. complex B to B^{act}). The stalled complex including the helicase is purified, and then transferred into micrococcal nuclease treated extract in which a second incubation initiates remodelling which is captured by UV irradiation over time. The tethered protein-RNA complexes are subsequently applied to the regular *in vitro* spliceosome iCLIP procedure as described in **Figure 6-1**. (**right**) Potential workflow for time resolved *in vitro* spliceosome iCLIP starting from purified spliceosomes, which were stalled by depleting the target helicase from splicing extract. The recombinantly produced helicase is then loaded onto the purified spliceosome before the chase is initiated by addition of the cofactors NTP and Mg^{2+}. A protein trap such as poly-U can ensure single turnover conditions. Additional recombinant factors can be added to follow an extended remodelling event. UV irradiation over time and the subsequent general iCLIP procedure (compare **Figure 6-1**) will produce the time-resolved binding profiles. Ideally, the epitope tags of depleted and repleted helicase differ.

In **chapter 5**, these pilot experiments are described only as a small step towards time-resolved *in vitro* spliceosome iCLIP. For instance, finding the right time-scale will be one of the most challenging tasks. For Brr2 unwinding, the literature gives an approximate duration of 4 sec at 4 °C (Small et al.

2006), a time frame which would be too short for the current crosslinking format. Hopefully, the implementation of slow- or non-hydrolysable NTPs, low concentrations of Mg^{2+} and low temperatures will increase the apparent reaction time. To ensure precise timing and mixing of components, kinetic experiments could be performed in a microfluidic setup which leads the reactants in a controlled manner directly to the UV light source. Similar innovations are in the development phase for time-resolved cryoEM (Frank 2017).

By comparing the binding profile of helicases on different splicing substrates, time-resolved *in vitro* spliceosome iCLIP, once developed, will allow us to draw many mechanistic conclusions. In the future, binding profiles of the different spliceosomal helicases might enable us to follow the entire spliceosome remodelling process from assembly and activation over catalysis to disassembly from the viewpoint of the spliceosomal helicases.

6.2 Implications of *in vitro* spliceosome iCLIP for the understanding of spliceosomal helicase functions

As described in the previous section, *in vitro* spliceosome iCLIP can be adapted to reveal the binding profile of additional spliceosomal helicases. We showed that Prp22 is a suitable target for *in vitro* spliceosome iCLIP in both complex C* and P (chapters **5.1** and **5.2**).

Being straightforward to adapt to further targets, we predict *in vitro* spliceosome iCLIP and time resolved *in vitro* spliceosome iCLIP will prove valuable for the understanding of spliceosomal helicases. First some general aspects are listed, while implications for answering some more specific questions are described in further detail afterwards.

(1) To date, we still lack knowledge about the RNA targets for some spliceosomal helicases which could be identified through binding profiles.

(2) It is unclear, how and when helicases are positioned and recognise a specific spliceosome stage. The time-point when they are released is equally unknown.

(3) Is the winching mechanism a universal model for all spliceosomal DEAH-box helicases?

(4) How processive are spliceosomal helicases? It is unclear if only one helicase molecule is involved per remodelling event and wheather consecutive release and rebinding of RNA by one or several helicase molecules is required.

(5) Proofreading activity of helicases can potentially be tested with suboptimal splicing substrates in *in vitro* spliceosome iCLIP and thereby identify the RNA target region during proofreading.

6.2.1 Towards a better understanding of Prp16 helicase

The role of Prp16 in 3'SS selection could, for example, be directly investigated by *in vitro* spliceosome iCLIP using a splicing substrate harbouring two 3'SS consensus sequences. It was reported that the splice site which is proximal to the brA is favoured over the distal site (Luukkonen and Seraphin 1997), but it remains unclear how and if Prp16 has an influence in the choice. The translocation path of Prp16 generated from splicing substrates with a single 3'SS and from splicing substrates with varied distances between two 3'SS might help to shed light on the mechanism. Additionally, sequencing approaches have the advantage of determining the usage of splice sites in a quantitative manner (Burke et al. 2018).

Prp16 was proposed to proofread the branching step by rejecting pre-mRNA substrates with branch point mutations (Burgess and Guthrie 1993; Koodathingal et al. 2010). By obtaining Prp16's binding profile in the context of a suboptimal BPS, and by comparing binding profiles of Prp16 on substrates with two branch point sequences, we might gain a better understanding of Prp16 mediated proofreading.

6.2.2 Towards a better understanding of Prp22 helicase

Prp22 is one of the best-studied spliceosomal helicases. The RNA binding position was determined by 4-thiouridine photo crosslinking to include the 17th nucleotide of exon 2 when mRNA release was blocked with dominant negative Prp22 (Schwer 2008). In agreement with the cryoEM structure of complex P (Bai et al. 2017; Liu et al. 2017; Wilkinson et al. 2017), this should be the starting point for Prp22-mediated mRNA release. Little is known, however, about the mechanism. Time-resolved *in vitro* spliceosome iCLIP might allow us to determine if Prp22 releases the RNA for example in a corkscrew-like fashion or if a short translocation on the mRNA is enough to destabilise interactions to U5 snRNA and proteins, which then leads indirectly to mRNA release.

In addition to its conventional role in mRNA release, Prp22 was reported to bind close to the 3'SS within the last eight nucleotides of the intron before the second catalytic step (McPheeters, Schwer, and Muhlenkamp 2000; McPheeters and Muhlenkamp 2003). The studies were done on spliceosomes which stalled due to a 3'SS alteration, however it is unclear which exact spliceosomal state was captured. Generating *in vitro* spliceosome iCLIP data for Prp22 from various spliceosomal stages might answer the question. In the cryoEM structure of complex C*, the distance between the catalytic centre and Prp22 is claimed to equal 16 or 17 nucleotides (Fica et al. 2017), which the authors put into agreement with the finding of (Schwer 2008), despite a low local resolution which does not allow the RNA path to be visualised.

After having characterised this second binding site of Prp22, there might be the possibility to reconstruct the complete transition towards exon ligation and mRNA release to analyse how Prp22

changes between the two binding locations lying in opposite direction to its translocation specificity. The intronic position, could of course also derive from Prp22 proofreading activity of the 3'SS, supported by the fact that dysfunctional Prp22 permits splicing of substrates with 3'SS mutations (Mayas, Maita, and Staley 2006).

6.2.3 Towards a better understanding of Brr2 helicase

As described above, time-resolved *in vitro* spliceosome iCLIP is a suitable tool to study Brr2-mediated U4/U6 snRNA duplex unwinding in the natural context of the spliceosome. Components, which join the spliceosome during the complex B to B^{act} transition, can be supplied from micrococcal nuclease deactivated extract (**Figure 5-7**).

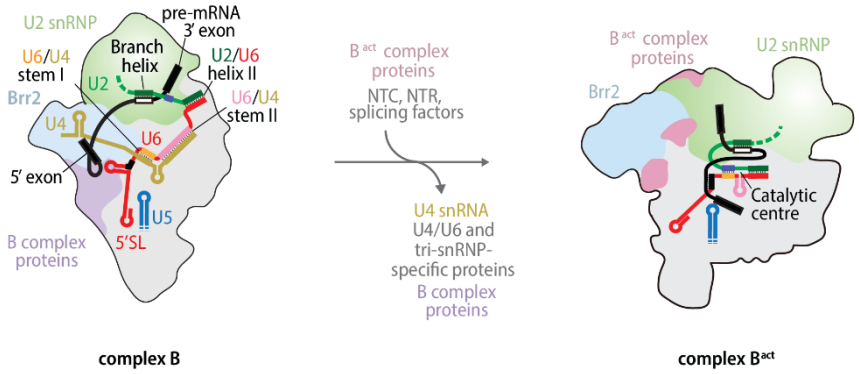


Figure 6-3: Schematic illustration of the RNA interaction network in complex B and B^{act}. In complex B, Brr2 (light cyan) is pre-loaded onto U4 snRNA (yellow). To unwind the U4/U6 snRNA duplex, Brr2 unwinds stem I first and then faces the U4 5'SL before unwinding of stem II can happen. Illustrations are adapted from (Plaschka, Lin, and Nagai 2017).

Following the path of Brr2 in a time-resolved manner would be an excellent approach to understand in which order specific RNA structures are actively remodelled by Brr2 in order to allow formation of the catalytic centre; bridging the data from *in vivo* studies in (Hahn et al. 2012) and the recombinant system in (Mozaffari-Jovin et al. 2012).

One of the main open questions can possibly be answered: After Brr2 has left its preloaded position on U4 snRNA and unwound U6/U4 stem I of the duplex, what happens when the helicase encounters the U4 5' stem loop (5'SL) which is tightly held by proteins in the centre of complex B (compare **Figure 6-3**)? Four possible scenarios were described in (Nielsen and Staley 2012): (1) Brr2 could potentially unwind U4 5'SL, while using an RNPase activity to replace the proteins and then continue to unwind stem II (Laggerbauer, Achsel, and Luhrmann 1998; Mozaffari-Jovin et al. 2012). (2) The active unwinding of the U4 5'SL and protein replacement by Brr2 could have a destabilising effect onto stem II which separates indirectly. This model can be supported by *in vivo* data, in which Brr2 crosslinks to the 3'-side of the U4 5'SL but not to stem II (Hahn et al. 2012). (3) Brr2 could

directly continue unwinding stem II after stem I without encountering U4 5'SL. (4) The last model suggests Brr2 would unwind only stem I, and then stop before the 5'SL. The unwinding of stem II would then actively be done by other factors (Theuser et al. 2016).
In addition to generating the binding profile to conclude where exactly Brr2 actively translocates, one could implement hiCLIP, which would enable the identification of duplex RNA regions in the vicinity of the crosslinked protein. Pinpointing what determines the timing of U4/U6 snRNA unwinding, an event that certainly has to be regulated accurately to avoid unwinding before complete spliceosome assembly, will be challenging and requires the interplay between different techniques.

Beside investigation of the canonical role for Brr2 in spliceosome activation, the helicase was proposed to be involved prior and at the second step of splicing (Hahn et al. 2012). Most of the results in this study were based on Brr2-G858R, which carries a mutation in the β-hairpin motif that suppresses defects in exon ligation without hindering the first step of splicing. *In vitro* spliceosome iCLIP could produce binding profiles for Brr2 to identify a direct RNA target in the later spliceosomal stages.

6.2.4 Towards a better understanding of spliceosome disassembly

Two helicases have been proposed to be involved in disassembly of the ILS. The target of the DEAH-box helicase Prp43 has not yet been identified unambiguously. Both the disruption of the U2 snRNA-BPS helix (Fourmann et al. 2016) and translocation on U6 snRNA (Wan et al. 2017) were suggested as Prp43 targets during disassembly. A combined approach of Prp43 *in vitro* spliceosome iCLIP in ILS stalled by dominant negative Prp43 and genome-wide iCLIP data from Prp43 *in vivo* should help to identify the helicase substrate.
In addition, Prp43 is proposed to disassemble spliceosomes that were directed to the discard pathway by Prp16 or Prp22 (Koodathingal et al. 2010; Mayas et al. 2010). The RNA target of Prp43 in these particles could be analysed in the context of ΔPrp16 or ΔPrp22 spliceosomes by *in vitro* spliceosome iCLIP. How Prp43 is recruited to spliceosomes, however, is beyond the scope of *in vitro* spliceosome iCLIP.
Similarly, the role and RNA target of Brr2 during disassembly is unclear. *In vitro* spliceosome iCLIP on stalled ILS might allow one to determine if and where Brr2 is loaded in ILS. Time-resolved *in vitro* spliceosome iCLIP of ΔPrp43 ILS, in which disassembly is induced by recombinant Prp43 would improve our understanding of disassembly.

6.3 Spliceosome remodelling and future directions

In vitro spliceosome iCLIP is of course not restricted to obtain binding profiles for the translocating helicases. Similarly, the tool can be applied to investigate DEAD-box helicases. Possibly, we can learn more about their duplex melting activity (Cordin and Beggs 2013), mainly when the investigation is combined with the hiCLIP approach. Further, binding profiles for splicing factors such as Slu7 and Prp18 could be targeted in the future. Maybe we can learn why some splicing factors are specifically required for substrates with longer distances between the brA and the 3'SS (James, Turner, and Schwer 2002; Ohrt et al. 2013).

Further, we predict that *in vitro* spliceosome iCLIP can easily be transferred from the intron reduced yeast to the human system. All data obtained *in vitro* could be complemented with either *in vivo* iCLIP or with spliceosome profiling in order to test gained hypotheses in a genome-wide environment and to increase our knowledge of substrate-specific effects.

Many open questions regarding spliceosome remodelling, however, are beyond the scope of *in vitro* spliceosome iCLIP and require a combination of different approaches. Conceivably, single molecule studies, time-resolved cryoEM or even tomography of the nucleus will help to answer the open questions:

(1) How are suboptimal substrates recognised during proofreading? What promotes the reaction to proceed backwards and into a discard pathway to increase splicing fidelity?
(2) Which parts of spliceosome remodelling require active restructuring and what can be considered as passive remodelling?
(3) How is the consecutive action of the helicases coordinated? What determines when each helicase joins or leaves the spliceosome? How do helicases work together in the way predicted for Brr2 and Prp16?
(4) How do helicases relocate when they are involved at different stages of splicing for different functions?

Most likely, many more questions will arise during the journey until we fully understand spliceosome remodelling which is essential to drive the splicing process, while providing regulation- and checkpoints. Finally, all this will bring us a step closer to understanding the process of eukaryotic gene expression.

7 Materials and methods

RNA and protein sequences used throughout the study derived from the yeast *Saccharomyces cerevisiae* unless otherwise stated.

7.1 Plasmids and yeast strains

7.1.1 Plasmids for *in vitro* RNA transcription

In general, DNA sequences were inserted under the T7 promoter into pUC cloning vectors for *in vitro* transcription. The modified U5 snRNA A, B and C sequences were PCR amplified from a plasmid containing full length U5 snRNA (provided by W. Galej) and cloned into pUC18 by C Oubridge. Additionally, C Oubridge inserted a 5'-hammerhead ribozyme sequence (Price et al. 1995) upstream of the U5 snRNA sequence.

Plasmid 1 was a gift from A. Newman, containing the sequence of three MS2 stem loops (Zhou, Sim, et al. 2002), the last 25 nt of the first exon of *UBC4*, the 84 nt-long *UBC4* intron, and the hepatitis delta ribozyme under the T7 promoter.

Plasmid 2 was designed to transcribe *UBC4*_3xMS2. The sequence for the transcription of the three MS2 stem loops was firstly cloned into the BamHI and EcoRI site of pUC19 by assembly of four DNA oligonucleotides (**Table 7-1**, primers 1 to 4). The oligonucleotides were enzymatically phosphorylated by T4 PNK (NEB) and annealed by decreasing the temperature slowly from 80 to 23 °C before ligation. In a second step, a construct containing the T7 promoter, the last 25 nt of the first exon, the intronic sequence and 25 nt of the second exon of *UBC4* were PCR amplified from a plasmid (gift from S. Fica) with primers compatible for HindIII- and BamHI-based cloning into the vector containing the 3xMS2 sequence (**Table 7-1**, primers 5 and 6).

7.1.2 Plasmids for protein tagging in *S. cerevisiae*

Vectors to introduce C-terminal His_8- and 3xFLAG-His_8-tag originated on pAG25 (Addgene plasmid #35121), which codes for the nourseothricin (ClonNAT) selection marker cassette downstream of a multicloning site. To generate pAG25-His_8, primers were constructed to contain the octa histidine-tag sequence (Tagwerker et al. 2006) while amplifying the 3' untranslated region (3'UTR) of

SV40 by PCR from a pRS424 expression vector (**Table 7-1**, primer 7 and 8). Restriction digestion of the amplified DNA allowed subsequent cloning into the HindIII and BamHI sites of pAG25.

To generate pAG25-3xFLAG-His_8, the His_8-SV40-region was amplified form pAG25-His_8 extending the sequence by a primer coding for 3xFLAG (**Table 7-1**, primer 8 and 9) and cloned into the HindIII/BamHI sites of pAG25.

Table 7-1: Oligonucleotides for plasmid cloning to be used in *in vitro* transcription and endogenous protein tagging. Primers are listed under the number as they appear in the text, their name and the DNA sequence.

#	oligonucleotide
1	primer 1 – MS2 5′-GATCCGATATCCGTACACCATCAGGGTACGAGCTAGCCCATGGCGTACACCATCAGGGTACGA-3′
2	primer 2 – MS2 5′-CTAGTAGATCTCGTACACCATCAGGGTACGG-3′
3	primer 3 – MS2 5′-CATGGGCTAGCTCGTACCCTGATGGTGTACGGATATCG-3′
4	primer 4 – MS2 5′-AATTCCGTACCCTGATGGTGTACGAGATCTACTAGTCGTACCCTGATGGTGTACGC-3′
5	HindIII – T7 promoter – UBC4 (forward) 5′-TCGAAAGCTTTAATACGACTCACTATAGGGAACTAAGTGATCTAGAAAGG-3′
6	BamHI – UBC4 (reverse) 5′-AGCTTGGATCCAACATGAAGTAGGTGGATCTCTAGTTCAATAG-3′
7	HindIII – His_8 – SV40 3′UTR start (forward) 5′-GTGCCTCGAAGCTTCAGGGGTTCACATCATCACCACCATCATCACCACTAATAGTAAGTCGACTTTGTTCCCACTG-3′
8	BamHI – SV40 3′UTR end (reverse) 5′-GTTTAAGAGGATCCGATCTTCGAGGGGAATTCTC-3′
9	HindIII – 3xFLAG – His (forward) 5′-GTGCCTCGAAGCTTCGACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGATTACAAGGATGACGATGACAAGAGGGGTTCACATCATCAC-3′

7.1.3 Yeast strains

The yeast strain BCY123 (*MATa*, *can1*, *ade2*, *trp1*, *Ura3-52*, *his3*, *leu2-3*, *112*, *pep4::his*$^+$, *prb1::leu2*$^+$, *bar1::HisG*$^+$, *lys2::pGAL1/10-GAL4*$^+$) is the mother strain for all modifications unless stated otherwise.

7.1.4 Protein tagging in *S. cerevisiae* strains

To introduce the sequence coding for a C-terminal protein tag into the yeast genome, the region containing the tag, the 3'UTR of SV40 (terminator) and the ClonNAT resistance cassette were amplified from pAG25-His_8 or pAG25-3xFLAG-His_8. Herby, the forward primer was designed with ~60 nt homology to the target gene upstream of the stop codon and 20 nt sequence similarity to the vector for amplification. The reverse primer was designed accordingly with 60 nt homology to the beginning of the 3'UTR of the target gene (compare **Table 7-2**).

The PCR product was gel purified (QIAEX II Gel Extraction Kit, Qiagen) and concentrated by precipitation to allow chemical transformation of at least 100 ng DNA into 20 µL BCY123 cells (lithium acetate method). Cells were kept in 1 mL yeast extract peptone dextrose (YEPD) for at least 9 h at 30 °C before further growth on selective plates containing 100 µg/mL ClonNAT. After three days, the genomic DNA of ClonNAT-positive cells was extracted (lithium-SDS method) and analysed for the insertion by PCR amplification with primer pairs annealing 300 nt up- and downstream of the original stop codon of the target gene. Successfully amplified insertions were verified by Sanger sequencing.

Additionally, western blot probing with anti-FLAG® M2-Peroxidase (Sigma) for the 3xFLAG-tag verified expression of the correctly tagged protein. C. Norman conducted most parts to generate Brr2-3xFLAG-His_8 strain.

Table 7-2: Oligonucleotides for endogenous protein tagging by homologous recombination. The first line shows the forward primer used to introduce the respective sequence downstream of the coding region of the target gene, the second line gives the reverse primer. Prp16-His_8 and Prp16-3xFLAG-His_8 were generated by amplification with the same reverse primer from the two different vectors.

tagged protein	DNA oligo sequence
Prp16-His_8	5'-GCAAAATATACTGAACGGCAAAGAAAATTCAATGAAACCTTTCAAAAGAAGGAAGCCTTTTTTTAGGGGTTCACATCATCACCAC-3'
	5'-GCATGCATATAACTATATAATAACATATATGAATATTTTGCCTATTAGCACGCTCTTCCCATAAAGCGGCCGCATAGGCCACTAGTGG-3'
Prp16-3xFLAG-His_8	5'-GCAAAATATACTGAACGGCAAAGAAAATTCAATGAAACCTTTCAAAAGAAGGAAGCCTTTTTTTGACTACAAAGACCATGACGGTG-3
Prp22-3xFLAG-His_8	5'-GGATCAAAATTCATGGAGACTAAGCTCAATAAGGCAGTCAAGGGAAAGGGCATTAGGTATCAAGAGGGACTACAAAGACCATGACGGTG-3'
	5'-GTTGTTAAAAAATTAAATATAGGTCTATAAAACTCGATAATTATAATGCATAAAAAGCTAACAATGGCGGCCGCATAGGCCACTAGTGG-3'
SmB-3xFLAG-His_8	5'-GTTTAATAATGAAGCGCCCCCTCAAACAAGGAAGTTTCAGCCCCCACCAGGTTTTAAAAGAAAAGACTACAAAGACCATGACGGTG-3'
	5'-CACATGCGTACACAAAAAAAGTATACGGAAACTATATTAGACTACACTACATCAACCTTAGCGGCCGCATAGGCCACTAGTGG-3'
Brr2-3xFLAG-His_8	5'-GGTGTGTCTGTGATTCATATCTTGACGCAGATAAAGAGTTGTCCTTTGAAATAAATGTGAAAGACTACAAAGACCATGACGGTG-3'
	5'-GCCGCAAGAGAATGTTATATATTGAAATCCATTCGATTATCCAGGACTAAACAATGATTGCGGCCGCATAGGCCACTAGTGG-3'

7.1.5 Plasmid shuffle

Shuffling plasmids expressed the Snu114 gene under its own promoter in pRS413 vector, which has a centromeric replication origin and codes for the resistance marker *HIS3* (gift from A. Newman). Mutations were introduced using the Kunkel mutagenesis method (Kunkel 1985) by either K. Nguyen or myself. The plasmids were transformed into the YSNU114KO1 strain (Frazer, Lovell, and O'Keefe 2009), in which the genomic copy of Snu114 is deleted. The essential Snu114 gene is provided on a pRS416 vector, which carries the *URA3* selection cassette and a centromeric replication origin. Transformants were first grown on selective plates lacking histidine and positive clones were transferred onto plates containing 5-FOA (5-fluoro orotic acid), which tests the growth phenotype after the Snu114-WT copy is lost under *URA3* anti-selection. Viable mutant clones were further transferred onto YEPD to monitor growth phenotypes at different temperatures.

7.1.6 Plasmids for recombinant protein expression

Vectors for recombinant expression in yeast were based on pRS424 and pRS426 vectors carrying the selection cassette for *TRP1* and *URA3*, respectively, to allow co-expression. The desired coding sequence is under control of the GAL-GAP promoter. Construct modification by Kunkel mutagenesis (Kunkel 1985) was verified by sequencing. Expression plasmids for helicases are listed together with mutagenesis primers in **Table 7-3**.

Table 7-3: Plasmids for recombinant expression of spliceosomal helicases in yeast. To allow expression in YM4 selective media, pRS424 or pRS426 (# 4 or 5) were co-transfected into yeast when expressing a single protein. Depending on the directionality of the GAL-GAP cassette, forward (for) or reverse (rev) primers were used in mutagenesis.

#	protein	N-term. tag	C-term. tag	vector	modification
1	Prp16	-	CBP-3xFLAG-His_6	pRS424	AA substitution on #2 to obtain WT
	primer (rev.)	5′-CATATAAATACTGTGCAAGTTGCGTGGTTTTACCTGAGCCCGTTTCACCAATTATCACC-3′			
2	Prp16-G378A	-	CBP-3xFLAG-His_6	pRS424	3xFLAG-tag insertion into Prp16-G378A-CBP-His (gift from A. Newman)
	primer (rev.)	5′-CTTTTTCCATCGTCGCTTCGATCCCCCCTTGTCATCGTCATCCTTGTAATCGATATCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCAAAAAAAGGCTTCCTTCTTTTG-3′			
3	Prp22-K512A	CBP	3xFLAG-His_6		3xFLAG-tag insertion into CBP-Prp22-K512A-His (gift from A. Newman)
	Primer (rev.)	5′-CCCTCAATGATGATGATGGTGATGATGTGAACCCCTCTTGTCATCGTCATCCTTGTAATCGATATCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCCCTCTTGATACCTAATGCCCTTTC-3′			
4	-	-	-	pRS424	Gift from A. Newman
5	-	-	-	pRS426	Gift from A. Newman

Plasmids for Snu114 and Prp8-N expression were constructed by C. Oubridge, W. Galej and myself (**Appendix Table 8-1** and **Table 8-2** indicating location of purification tags and expression success).

For bacterial expression vectors, constructs were PCR amplified from plasmids containing the coding sequence for Prp8 and cloned into either pGEX-6P-I or pET-based vectors by restriction-enzyme based cloning. In addition, a codon-optimised construct of Prp8 was synthesised directly by gene synthesis (Invitrogen) and similarly transferred into a pET-based expression vector. All constructs are listed in **Appendix Table 8-1**.

7.2 RNA techniques

Several different strategies were used to produce the RNA molecules included in this work. The following paragraphs will describe the general methods separately, while **Table 7-4** will specify which methods were used to generate each specific RNA transcript. **Table 7-5** lists reaction buffers and their components.

Table 7-4: RNA transcripts generated for this work. The name of the transcript is listed as and where (column chapter) it appears in the thesis. Details on how the molecule was generated are given under methods.

transcript	chapter	method
U5 snRNA A	2.1	*In vitro* transcription from plasmid
U5 snRNA B	2.1	*In vitro* transcription from plasmid
U5 snRNA C	2.1	*In vitro* transcription from plasmid
3xMS2_*UBC4*_short	4.1, 4.2, 4.3	*In vitro* transcription from plasmid 1, ligation to a Cy5-labelled 36 nt RNA oligonucleotide
3xMS2_*UBC4*_AC_short	3, 4.1, 4.2, 4.3	*In vitro* transcription from plasmid 1, ligation to a Cy5-labelled 36 nt RNA oligonucleotide
3xMS2_*UBC4*_dG_short	5.1	*In vitro* transcription from plasmid 1, ligation to a Cy5-labelled 36 nt RNA oligonucleotide
*UBC4*_3xMS2	5.2	*In vitro* transcription from plasmid 2, 3'-end labelling with fluorescein
3xMS2_*UBC4*_long	4.1.2	*In vitro* transcription from plasmid 1, two consecutive ligations to attach two 36 nt RNA oligonucleotides with the second carrying a Cy5 label
3xMS2_*UBC4*_AC_long	4.1.2	*In vitro* transcription from plasmid 1, two consecutive ligations to attach two 36 nt RNA oligonucleotides with the second carrying a Cy5 label
3xMS2_*ACT1*_WT	4.4	*In vitro* transcription of PCR generated template followed by enzymatic Cy5-labelling
3xMS2_*ACT1*_AC	4.4	*In vitro* transcription of PCR generated template followed by enzymatic Cy5-labelling

Table 7-5: **Buffers and reaction conditions used for RNA preparation.** The name of the buffer is listed together with its composition., The total volume is given in brackets for reactions which were conducted in a specific volume. Buffers were prepared with Milli Q water (Millipore) and filtered through 0.22 µm filters. Enzymatic reactions were set up with RNase free water (Ambion). Tris: tris(hydroxymethyl)aminomethane, DTT: dithiothreitol

buffer	composition
oxidation solution	0.1 M sodium peroxide, 0.1 M sodium acetate pH 5.3
Cy5 labelling reaction	1 nmol RNA, 1 mM ATP, 1 x RNA ligase buffer (NEB), 10 % dimethyl sulfoxide (DMSO), 2 nmol pCp-Cy5 (= 2-fold excess over RNA), 1 µL RNasin (Promega), 2 µL RNA ligase I (high concentration) (25 µL total)
in vitro transcription	0.1 mg/mL DNA template (for plasmid DNA; PCR product optimum usually lower), 24 µg/mL T7 RNA polymerase (purified by K. Nagai or C. Oubridge), 40 mM Tris-Cl pH 8.0, 15 mM $MgCl_2$, 10 mM DTT, 2 mM spermidine and 4 mM of each ATP, GTP, CTP and UTP
RNA elution buffer	300 mM sodium acetate pH 5.3, 0.1 % SDS, 1 mM ethylenediaminetetraacetic acid (EDTA)
10 x TBE	108 g Tris base, 9.3 g EDTA, 55g boric acid in 1 L (LMB media kitchen)
formamide loading dye	95 % formamide, 10 mM EDTA, pinch of bromophenol blue
ribozyme cleavage buffer	5 mM Tris-Cl pH 7.4, 5 mM NaCl, 25 mM $MgCl_2$
2 x ligation mix	27 µL water, 8 µL 10 x T4 DNA ligase buffer (NEB), 1 µL RNasin (Promega) and 4 µL T4 DNA ligase (gift from K Nagai), (40 µL total)

Table 7-6: **Primers used for PCR derived transcription templates and to extend RNA fragments.** Primers are listed under the number as they appear in the method section, together with their name and sequence. DNA oligonucleotides (1 – 4b) highlight the 3'SS in bold; RNA oligonucleotides (5 – 8) carried a Cy5-label on the 3'-end as indicated in the text, the (dG) represents a deoxynucleotide.

#	oligonucleotide	sequence
1	XmaI – T7 promoter (forward)	5'-CGGGCGATCCCGGGTTTAATACGACTCACTATAGGCGG-3'
2	SacI – exon2-164 nt (reverse)	5'-ACAATCTGGAGCTCAGCTTCATCACCAACGTAGG-3'
3a	3'SS – WT (forward)	5'-ATGTTT**AGAG**GTTGCTGCTTTGGTTATTGATAA-3'
3b	3'SS – ACAC (forward)	5'-ATGTTT**ACAC**GTTGCTGCTTTGGTTATTGATAA-3'
4a	3'SS – WT (reverse)	5'-TTATCAATAACCAAAGCAGCAAC**CTCT**AAACAT-3'
4b	3'SS – ACAC (reverse)	5'-TTATCAATAACCAAAGCAGCAAC**GTGT**AAACAT-3'
5	UBC4_AC exon2 (-11 to 25)	5'-TATTGAACTACACATCCACCTACTTCATGTTCACGC-3'
6	UBC4_dG exon2 (-11 to 25)	5'-TATTGAACTA(dG)ACATCCACCTACTTCATGTTCACGC-3'
7	UBC4 exon2 (-11 to 25)	5'-TATTGAACTAGACATCCACCTACTTCATGTTCACGC-3'
8	UBC4 exon2 (26 to 61)	5'-GGTCCACTCGGCGATGATCTATATCACTGGCAAGCA-3'

7.2.1 *In vitro* RNA transcription

Templates for *in vitro* transcribed RNA were either PCR product or plasmid DNA containing the desired sequence under the T7 promoter. Plasmid DNA was linearized by restriction digestion prior

to transcription. Transcription reactions were performed in volumes from 500 µL to 10 mL containing 0.1 mg/mL linearized plasmid. Optimal concentrations of PCR template input were determined by small scale transcription.

To generate a PCR fragment for the *ACT1* transcript, a region containing the T7 promoter, the last 25 nt of the first exon, the intron and 25 nt of the second exon were first PCR amplified (primers 1 and 4 a/b in **Table 7-6**) from a pUC-like vector (gene synthesis). Secondly, the sequence of *ACT1* starting 8 nt upstream of the 3'SS until nucleotide 164 of the second exon was amplified from genomic BCY123 DNA (primers 2 and 3a/b). Lastly, these two DNA fragments overlapping by 33 nt were annealed and amplified by PCR (primers 1 and 2). Mismatches in the primer sequences at the 3'SS allowed to generate both the WT and the AC-version of the transcript (named a and b in **Table 7-6**, the 3'SS sequence is highlighted in bold).

DNA, derived from either linearized plasmid or PCR fragment, was purified by PCI (phenol:chloroform:isoamyl alcohol, 25:24:1, v/v) extraction and precipitated before *in vitro* transcription for 4 h at 37 °C (reaction composition in **Table 7-5**). Precipitated magnesium phosphate was pelleted by centrifugation, and the supernatant purified by PCI extraction followed precipitation with either ethanol or isopropanol at -20 °C. RNA containing pellets were dissolved in formamide loading dye and separated on 1 x TBE 7 M urea PAGE (acrylamide percentage depending on the transcript length). The RNA product was visualised by UV shadowing and extracted by electroelution in 1 x TBE or by passive elution in RNA elution buffer. The extracted RNA was either concentrated using Amicon Utra centrifugal filter concentrators (Millipore) or PCI extracted and precipitated until further use. Final concentration was measured by absorption at 260 nm using a Nanodrop spectrophotometer (Thermo Scientific).

7.2.2 3'-end labelling of RNA with fluorescein

In vitro transcribed RNA, gained from 1 mL transcription reaction, was resolved in 500 µL oxidation solution and incubated for 1.5 h in the dark at 23 °C. The reaction was quenched with 0.25 M KCl (final concentration) and kept on ice for 10 min, before removal of the precipitate by centrifugation. The supernatant was incubated with 550 µL fluorescein-5-thiosemicarbazide (Sigma) for 4 h in the dark at 23 °C. The fluorescein labelled RNA was separated from free label and transferred into H_2O using a NAP-10 column (GE Healthcare) according to the manufacturer's instructions.

7.2.3 Enzymatic 3'-end labelling of RNA with Cy5

1 nmol *in vitro* transcribed RNA was enzymatically labelled by RNA ligase I (NEB, high concentration) in a 25 µL Cy5-labelling reaction for 1 h at 37 °C (**Table 7-5**). Excess pCp-Cy5 (Jena Bioscience) was

removed by phenol extraction. The labelled RNA was precipitated and dissolved in a suitable amount of H_2O.

7.2.4 Ligation of *in vitro* transcribed RNA and oligonucleotides

Some *UBC4* transcripts were partially produced by *in vitro* transcription and then extended by ligation to RNA oligonucleotides (Sigma or Integrated DNA Technologies, **Table 7-6**). First, the DNA sequence coding for 3xMS2 stem loops, exon 1 and parts of the intron of *UBC4* ending with the hepatitis delta virus ribozyme was transcribed *in vitro* in a 1 mL reaction (compare **7.2.1, plasmid 1**). After precipitation, the RNA was dissolved in 400 µL ribozyme cleavage buffer and cleavage induced by the temperature ramps given in brackets (2 cycles: 5 sec at 80 °C, 20 min temperature ramp to 50 °C; 10 min temperature ramp to 30 °C).

The product was separated from the primary transcript and the ribozyme by denaturing PAGE (1 x TBE, 7 M urea, 5 % polyacrylamide) and visualised by UV shadowing. RNA was eluted in 5 volumes RNA elution buffer at 37 °C for 14 h before PCI extraction and precipitation. The 2',3'-cyclic phosphate on the 3'-end was removed by 3 µL T4 PNK in 200 µL 1 x T4 PNK buffer (both NEB) at 37 °C for 120 min. The reaction was incubated for further 60 min after the addition of 10 µL 1 M Tris-Cl pH 8.0 and 1 µL AP (Roche, 20 U/µL). The final product was cleaned up by PCI extraction, precipitated and dissolved in H_2O. Final concentration was measured by absorption at 260 nm.

Second, one or two RNA oligonucleotides were attached to generate *UBC4* short and long transcripts, respectively. 2,000 pmol RNA oligonucleotides were phosphorylated in a 100 µL reaction containing 1 x T4 DNA ligase buffer (NEB), 1 µL RNasin and 4 µL T4 PNK for 90 min at 37 °C. This RNA oligonucleotide was ordered with a Cy-5 label at the 3'-end for the short versions (25 nt exon 2). The product was purified by phenol extraction, precipitated and dissolved in H_2O. 1,000 pmol each of the *in vitro* transcribed RNA fragment, the phosphorylated RNA oligonucleotide and a DNA oligonucleotide with at least 30 nt complementary sequence spanning both RNA fragments, were annealed in 40 µL H_2O by slowly decreasing the temperature from 80 °C to 25 °C. After addition of 40 µL 2 x ligation mix the reaction was incubated at 25 °C for 16 hours, PCI extracted and precipitated. The final product was separated from non-ligated components and the DNA oligonucleotide by denaturing PAGE (5.5 % polyacrylamide, 7 M urea, 1 x TBE), excised and eluted in elution buffer at 37 °C for 16 h. After PCI extraction and precipitation, the concentration was determined by absorbance at 260 nm in H_2O.

For *UBC4*_long, the ligation step was repeated with an additional phosphorylated RNA oligonucleotide containing nucleotide 26-61 of *UBC4* exon 2 and a Cy-5 labelled 3'-end (primer 8, **Table 7-6**).

Some of the RNA transcripts produced according to this method were produced by A. Newman.

7.3 Spliceosome and protein purification including biochemical techniques

Table 7-7: Buffers used for expression and purification of spliceosomes and spliceosomal proteins. The name of the buffer is listed as it appears in the text together with the components. Buffers which were provided by the LMB media kitchen are marked as such. Buffers were prepared with Milli Q water (Millipore) and filtered through 0.22 µm. PEG: polyethylene glycol, ACN: acetonitrile, HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

buffer	Composition
YEPD	11 g yeast extract, 22 g peptone, 55 mg adenine sulphate, 100 mL 20 % glucose for final volume of 1 L (LMB media kitchen)
YM4	6.9 g yeast nitrogen base, 11 g casamino acid, 55 mg adenine, 55 mg tyrosine for a final volume of 1 L (LMB media kitchen)
2 x lysis buffer	2 M NaCl, 100 mM Tris-Cl pH 9.0, 2 mM imidazole, 20 mM β-mercaptoethanol (β-ME), 0.2 % IGE pal CA-630, 4 mM $CaCl_2$, 2 mM magnesium acetate, EDTA-free protease inhibitor cocktail (Roche)
CAL wash buffer	500 mM NaCl, 20 mM Tris-Cl pH 8.0, 2 mM $CaCl_2$, 1 mM magnesium acetate, 1 mM imidazole, 10 mM β-ME
CAL elution buffer	500 mM NaCl, 20 mM Tris-Cl pH 8.0, 2 mM EGTA, 1 mM magnesium acetate, 1 mM imidazole, 10 mM β-ME
Ni-NTA binding buffer	1 M NaCl, 20 mM Tris-Cl pH 8.5, 5 mM imidazole, 10 mM β-ME
Ni-NTA wash buffer	1 M NaCl, 20 mM Tris-Cl pH 8.5, 15 mM imidazole, 10 mM β-ME
Ni-NTA elution buffer	1 M NaCl, 20 mM Tris-Cl pH 8.5, 250 mM imidazole, 10 mM β-ME
storage buffer Prp16	250 mM KCl, 20 mM HEPES pH 7.9, 0.2 mM EDTA, 0.5 mM DTT, 20 % glycerol (similar to buffer D)
storage buffer Prp22	300 mM KCl, 20 mM HEPES pH 7.9, 0.2 mM EDTA, 0.5 mM DTT, 20 % glycerol (similar to buffer D)
storage buffer Prp8/Snu114	500 mM NaCl, 20 mM Tris-Cl pH 8.5, 5% glycerol, 1 mM DTT
TEV buffer	1M NaCl, 20 mM Tris-Cl pH 8.5, 5 mM imodazole, 1 mM DTT
low salt buffer	500 mM KCl, 20 mM HEPES pH 7.8
reverse phase buffer	100 mM potassium phosphate pH 6.5, 10 mM tetra-n-butylammonium bromide, 7.5 % ACN
AKG buffer	10 mM HEPES pH 7.9, 1.5 mM $MgCl_2$, 200 mM KCl, 10 % glycerol, 0.5 mM DTT
buffer D	20 mM HEPES pH 7.9, 0.2 mM EDTA, 50 mM KCl, 20 % glycerol, 0.5 mM DTT
stop mix	1 mg/mL proteinase K, 50 mM EDTA, 1 % SDS
splice diluent	300 mM sodium acetate pH 5.3, 1 mM EDTA, 0.1 % SDS, 1 % glycogen
MS2-assembly buffer	3 % PEG6000, 60 mM KPO_4, 40 % buffer D
splice buffer	2 mM ATP, 2.5 mM $MgCl_2$, 60 mM potassium phosphate pH 6.5
G70 buffer	20 mM HEPES pH 7.9, 70 mM KCl, 0.1 % Igepal CA-630, 0.5 mM EDTA
K75	20 mM HEPES pH 7.9, 75 mM KCl, 0.05 % Igepal CA-630, 0.25 mM EDTA, 5 % glycerol

K75G10	20 mM HEPES pH 7.9, 75 mM KCl, 0.05 % Igepal CA-630, 0.25 mM EDTA, 10 % glycerol
maltose elution buffer	20 mM HEPES pH 7.9, 75 mM KCl, 0.05 % Igepal CA-630, 0.25 mM EDTA, 5 % glycerol, 12 mM maltose

7.3.1 Recombinant protein expression in yeast

BCY123 cells were co-transfected with expression vectors pRS426 and pRS424. Positive transformants were grown in 24 L YM4 selective media supplemented with 1 % raffinose at 30 °C. Protein expression was induced with 2 % galactose (final concentration) at OD_{600} 1.0. After 12-16 hours further growth at 30 °C, cells were harvested, resuspended in 1 volume 2 x lysis buffer, and frozen in liquid nitrogen in droplet form. Cells were disrupted in a 6870 Freezer/Mill (SPEX SamplePrep) under the following settings: 4 cycles, 2 min precool time, 2 min run time, 1 min cool time between cycles and a rate of 12 cps. After thawing, the pH of the extract was raised to 8.5 by Tris base. Cell debris were removed by ultracentrifugation at 195,000 x *g* for 90 min.

The supernatant was incubated with 2 mL Calmodulin-sepharose beads (produced by C. Oubridge) for 12-16 hours at 4 °C. Beads were washed with 5 x 50 mL CAL wash buffer and the proteins eluted in 10 x 5 mL CAL elution buffer. Protein-containing fractions were pooled and dialysed against Ni-NTA binding buffer for 4 h at 4 °C. After 14 h binding to 4 mL Ni-NTA agarose beads (Quiagen), the beads were first washed with 15 mL Ni-NTA binding buffer and then with 15 mL Ni-NTA wash buffer. The protein was eluted in about six 4 mL fractions with Ni-NTA elution buffer.

rPrp16-G378A, rPrp16 WT and rPrp22-K512A were dialysed against the respective helicase storage buffer for 4 h at 4 °C and stored at -80 °C until further use (see **7.3.10 splicing reactions**). For storage and removal of the purification tag of Snu114:Prp8-N dimers, the sample was incubated with TEV protease while dialysing against storage buffer for 14 h at 4 °C before storing at -80 °C. Samples containing a protein with an internal TEV protease cleavage site were equally incubated with TEV protease but dialysed against TEV buffer. Additional incubation with 4 mL Ni-NTA agarose beads was used to remove the cleaved and polyhistidine-tagged fraction including TEV protease, retaining the purified sample in the flow-through. Snu114:Prp8-N dimers were concentrated in an Amicon Ultra centrifugation filter concentrator at 16 °C while increasing the salt concentration to 0.9 M KCl by addition of concentrated KCl or buffer exchange. Subsequent gel filtration was performed in 1 M KCl, 20 mM Tris-Cl pH 8.5 and 1 mM DTT using a Superdex 200 10/300 gel filtration column (GE Healthcare). Protein-containing fractions were pooled and directly used for downstream applications.

7.3.2 Protein-RNA complex formation

Snu114:Prp8-N and RNA were mixed in a 1.25:1 ratio for reconstitution. The RNA was prepared in a buffer without salt (20 mM Tris-Cl pH 8.5, 10 mM β-ME) and added to the protein solution containing 1 M KCl to yield a final concentration of 0.2 M KCl. Fast mixing avoided exposure of the protein to low salt concentrations and therefore prevented aggregation. After incubation for 1 hour on ice, the sample was concentrated while simultaneously increasing salt to 0.3 M KCl. EMSA comparing free RNA and complex confirmed the efficiency of reconstitution (0.5 % agarose, 1 x TB, 40 V, 90 min; stained after electrophoresis with ~0.5 µg/mL ethidium bromide).

7.3.3 Limited proteolysis

Limited proteolysis was performed on Snu114:Prp8-N reconstituted with U5 snRNA (A) with three different constructs for Prp8-N (Prp8 122-735, Prp8 1-770, Prp8 122-770) (compare **7.3.2** Protein-RNA complex formation). In addition, Snu114:Prp8 1-770 was proteolyzed as a protein complex. Five different proteases were used: trypsin, chymotrypsin, subtilisin, Glu-C, and Lys-N. Proteolysis was performed with 5 µg protein at two different protease ratios (1:100 and 1:250 protein/protein-RNA-complex:protease) in a 20 µL reaction volume in either protein-RNA complex buffer (0.3 M KCl, 20 mM Tris-Cl pH 8.5, 1 mM DTT) or protein buffer (1 M KCl, 20 mM Tris-Cl, 1 mM DTT). Reactions were stopped after 30 min, 1 hour or 2 hours, respectively, by adding 7 µL of 2 x NuPAGE™ LDS Sample buffer (Invitrogen). Samples were heated at 90 °C for 3 min before separation on on 4-12 % NuPAGE Bis-Tris protein gels (Invitrogen).

7.3.4 Crystallography

Crystallisation trials were set up using the sitting-drop vapour diffusion technique in a 96 well format provided by the LMB crystallisation facility (**Table 7-1**) at both 277 K and 293 K. The protein solution (approx. 4.5 mg/mL in either (a) 300 mM KCl, 20 mM K-HEPES, pH 8.5, 1 mM DTT or (b) 0.2 M KCl, 0.15 M ammonium acetate, 20 mM Tris-Cl pH 8.5, 1 mM DTT) was mixed with an equal volume of reservoir solution to obtain 200 nL drops. For promising crystallisation conditions, four-corner screens varying two of the different components were used.

Table 7-8: LMB screening plates, used in this work, giving their composition of commercial screens.

LMB screen	commercial screens included in particular LMB screen
LMB1	Crystal Screen 1 (1-48) (Hampton Research)
	Crystal Screen 2 (1-48) (Hampton Research)
LMB2	Wizard 1 (1-48) (Emerald BioStructures)
	Wizard 2 (1-48) (Emerald BioStructures)
LMB4	PEG6000 (1-24) (Hampton Research)
	MPD (1-24) (Hampton Research)

	MembFac (1-48) (Hampton Research)
LMB5	PEG-Ion screen (1-48) (Hampton Research)
	Natrix (1-48) (Hampton Research)
LMB6	CrystalScreen Lite (1-48) (Hampton Research)
	CrystalScreen Cryo (1-48) (Hampton Research)
LMB8	JBS1 (PEG 400-3K) (JenaBioScience)
	JBS2 (PEG 4K) (JenaBioScience)
	JBS3 (PEG 4K +) (JenaBioScience)
	JBS4 (PEG 6K-8K) (JenaBioScience
LMB9	JBS 5 (PEG 8K-20K) (JenaBioScience)
	JBS 6 (Ammoniumsulfate) (JenaBioScience)
	JBS 7 (MPD) (JenaBioScience)
	JBS 8 (MPD/alcohols) (JenaBioScience
LMB10	JBS9 (alcohols/salt) (JenaBioScience)
	JBS10 (salt) (JenaBioScience)
	Clear Strategy Screen 1 pH 4.5 (1-24) Molecular Dimensions Limited (see below !!)
	Clear Strategy Screen 1 pH 5.5 (1-24) Molecular Dimensions Limited
LMB12	Clear Strategy Screen 2 pH 5.5 (1-24) Molecular Dimensions Limited
	Clear Strategy Screen 2 pH 6.5 (1-24) Molecular Dimensions Limited
	Clear Strategy Screen 2 pH 7.5 (1-24) Molecular Dimensions Limited
	Clear Strategy Screen 2 pH 8.5 (1-24) Molecular Dimensions Limited
LMB16	JCSG+ (Qiagen / Nextal)
LMB17	Morpheus Screen (F. Gorrec)

7.3.5 Nucleotide analysis by HPLC

Snu114:Prp8-N was prepared as described in **7.3.1** and concentrated to ~7.5 mg/mL. About 100 µL protein solution were dialysed against low salt buffer in a Slide-A-Lyzer MINI dialysis unit (Thermo Scientific, 10K MWCO) at 4 °C for 14 hours. To release the nucleotide, the protein was denatured at 90 °C for 5 min. 30 µL protein were diluted with 90 µL H_2O to decrease the salt concentration. The solution was mixed by vortexing and any precipitate removed by centrifugation. HPLC was run on a SUPELCOSIL™ LC-18-DB HPLC column (5 µm particle size, L × I.D. 25 cm × 4.6 mm, Supelco) with reverse phase buffer. GTP and GDP nucleotides were diluted to 25 µM in 1/4th low salt buffer and analysed equally for their retention volume.

7.3.6 Fluorescence polarisation

The concentration of U_{15}-RNA with a 6-carboxyfluorescein at the 5'-end (Integrated DNA Technologies) was held constant at 20 nM throughout the experiments. rPrp16 was mixed with the RNA substrate in 100 mM KCl reaction buffer in a 1.5-fold dilution series. Supplements of ATP and Mg^{2+} as well as the exact protein concentration are indicated in the respective figures (compare **chapter 5.4.2**). Fluorescence polarisation (FP) and fluorescence intensity (FI) were measured in a PHERAstar plate reader with the modules FP 485 520/520 and FI 485 520, respectively. The obtained data points were inspected for constant FI, before FP was blotted against the protein concentration. FP

data points were fit to a binding curve in GraphPad with the parameters for 'one-site – specific binding', and K_d values determined. One site – specific binding: $Y = B_{max}*X/(K_d + X)$; B_{max} = maximum specific binding; X in M; Y and B_{max} in relative units (RU).

7.3.7 Splicing extract preparation

Yeast cells were grown in YEPD at 30 °C. The cells were harvested by centrifugation, when reaching a density of 1.6 OD_{600}. Cell pellets were washed with H_2O and AKG buffer before resuspending in 0.6 volumes AKG buffer. Frozen cell droplets were disrupted in a 6870 Freezer/Mill (SPEX SamplePrep) with the following settings: 2 cycles, 2 min precool time, 2 min run time, 1 min cool time between cycles and a rate of 10 cps. Cell debris were separated by centrifugation at 37,000 x *g* for 30 min. Subsequent centrifugation at 117,000 x *g* for 75 min separated the extract into two layers. The top layer was dialysed twice against buffer D for 90 min at 4 °C. The dialysis time was extended to 2 h followed by 14 h for splicing extract for spliceosomal complex B purification. After removing precipitate during centrifugation at 3,000 x *g* for 5 min, 600 µL aliquots were frozen in liquid nitrogen and stored at -80 °C until further use.

7.3.8 *In vitro* splicing reactions

Small scale splicing reactions were performed with 2.5 nM fluorescently labelled pre-mRNA substrate and 40 % splicing extract in splice buffer and incubated at 23 °C. To obtain a time course, 10 – 20 µL samples were removed at various time points and mixed with 5 – 20 µL stop mix. The samples were deproteinised at 37 °C for 20 min. 100 µL splice diluent was added before PCI extraction. The aqueous phase was ethanol precipitated at -20 °C. Pellets were taken up in 4 – 7 µL formamide loading dye, dissolved by vortexing and heated up for 90 sec at 90 °C before separating on a denaturing PAGE (1 x TBE, 7 M urea, polyacrylamide percentage depending on the RNA substrate). Fluorescently-labelled RNA species were visualised with an Amersham Typhoon (GE Healthcare).

7.3.9 MS2-MBP assembly

90 nM MS2-tagged pre-mRNA substrate were heated up to 90 °C for 2 min in 75 µL MS2-assembly. Slowly decreasing the temperature to 23 °C ensured proper folding of the MS2 stem loops. $MgCl_2$ was added to a final concentration of 2 mM, and the reaction was incubated with 1.25 excess of MS2-MBP protein for 30 min at 4 °C. Recombinant MS2-MBP fusion protein was purified from *E. coli* by C. Oubridge.

7.3.10 First purification steps for spliceosomal complex C

Splicing of the MS2-MBP assembled pre-mRNA substrate was performed in a 1.5 mL reaction with 40 % splicing extract in splice buffer at 23 °C for 30 min. Splicing extract derived from the following yeast strains as described in the respective chapter: BCY123 (untagged), SmB-3xFLAG, Prp16-3xFLAG-His_8-tag, BCY123 supplemented with rPrp16-G378A. Recombinant rPrp16-G378A was pre-incubated at 4 °C in splicing extract at a concentration to yield 15 ng/mL in the splicing reaction. Pre-mRNA substrates are indicated likewise in the respective chapter.

After splicing, the reaction was adjusted to 2 mM glucose and the incubation prolonged for 5 min to deplete ATP. After dilution with 1.5 mL G70 buffer, precipitate was removed by centrifugation at 3,000 x *g*. The solution was layered onto 800 µL buffer G75 containing 40 % glycerol for ultra-centrifugation at 237,000 x *g* for 3 hours in a swing-out rotor. The high glycerol containing cushion was collected and adjusted to a final concentration of 0.025 % Igepal CA-630 before applying to 100 µL amylose resin (NEB). After nutation for 14 hours at 4 °C, the beads were washed thrice with K75 buffer. Spliceosomes were eluted with 200 µL maltose elution buffer. The elution efficiency was monitored in a PHERAstar plate reader with the modules FI 640 680 or FI 485 520 measuring the fluorescence intensity of the Cy5- or fluorescein-labelled RNA, respectively. Compare (Galej et al. 2016) for a similar purification strategy.

7.3.11 First purification steps for spliceosomal complex C*

Splicing was performed with the pre-mRNA substrate 3xMS2_*UBC4*_dG_short essentially as described for spliceosomal complex C. rPrp-K512A was used at a final concentration of 15 ng/mL. After binding on amylose, an additional step was inserted. The beads were washed with K75G10 and incubated with 2 mM ATP/2mM $MgCl_2$ in 1 mL K75G10 for 30 min nutating at 23 °C. Subsequent washes and the elution were performed according to the complex C procedure. This purification strategy was adapted from (Fica et al. 2017).

7.3.12 First purification steps for spliceosomal complex P

Splicing was performed as describes for complex C with *UBC4*_WT_3xMS2 pre-mRNA substrate and 15 ng/mL final concentration of rPrp-K512A. After splicing, the reaction was incubated with 5 µM of DNA oligonucleotide complementary to the 3'-exon (sequence: `ATGAAGTAGGTGGAT`) for another 20 min to induce cleavage of the 3'-MS2 tag by the RNase H activity present in the splicing extract. Spliceosomes without a docked 3'-exon were removed while complex P with a docked 3'-exon was protected from RNase H cleavage and could be subsequently purified via amylose capture as described for complex C. This purification strategy was adapted from (Wilkinson et al. 2017).

7.3.13 First purification steps spliceosomal complex B

Splicing extract was pre-treated with 2 mM glucose and incubated for 10 min at 23 °C. Spliceosomes were assembled onto 3xMS2_*UBC4*_WT as described for complex C, reducing ATP to 50 µM and extending the incubation time to 45 min.

In contrast to all other *in vitro* spliceosome iCLIP experiments described in this thesis, crosslinking for complex B was performed in a 6-well plate on the 1.5 mL splicing reaction with 312 µJ/cm^2 254 nm UV light (Stratalinker 2400). Purification of the complex via MS2-MBP fusion protein was done as for complex C. The purification strategy for complex B was inspired by (Plaschka, Lin, and Nagai 2017). The *in vitro* spliceosome iCLIP procedure continued with urea denaturation of the amylose elution (see **7.4.2**).

For chasing experiments (compare results **subchapter 5.3.3**), a 1.5 mL splicing reaction was set up with 40 % micrococcal nuclease treated extract, 2 mM ATP containing splice buffer, while the RNA proportion was replaced by purified complex B. 20 µL samples were taken at the indicated time points during incubation at 23 °C. A control was prepared in the same way with 50 µM ATP in the splice buffer. Glycerol cushion centrifugation was performed as described before.

7.3.14 Protein depletion from splicing extract via epitope tag

Prp16-3xFLAG depletion was done via its epitope tag using anti-FLAG M2 magnetic beads (Sigma). 100 µL splicing extract was incubated with 25 µL anti-FLAG beads equilibrated in TBS (50 mM Tris-Cl, 150 mM NaCl, pH 7.4), and nutated at 4 °C. After 1 hour, the extract was separated from the beads, and the whole procedure was repeated twice with 25 µL beads each time. After every step, 3 µL extract were probed for remaining Prp16-3xFLAG with anti-FLAG AB by western blot.

The depleted extract was tested for second step splicing defects in 20 µL splicing reactions. To rescue the phenotype, 2 µL of the splicing reaction were replaced by rPrp16-WT in buffer D ranging from 1.6 to 3.2 µg/mL final concentration in the reaction. In addition, controls supplemented with similar amounts of rPrp16-G378A or buffer D were tested for splicing.

7.3.15 Micrococcal nuclease treatment

100 µL splicing extract were incubated with 2 mM $CaCl_2$ and 2.5 µL micrococcal nuclease (NEB) for 30 min at 30 °C to remove endogenous RNA. The reaction was terminated by 8 mM EGTA and incubation for 1 min at 30 °C followed by 10 min at 4 °C. Treated extracts were directly used in splicing reactions.

For optimisations as shown in **subchapter 5.3.3**, 20 µL splicing extract were supplemented with 2 mM $CaCl_2$. To monitor nuclease efficiency, Cy5-lablled RNA was added to a 2.5 nM final concentration. To monitor the later inactivation, no RNA was added. The samples were incubated with 5, 10 or 20 units micrococcal nuclease per µL splicing extract and incubated at 30 °C for 30 min. The reaction was adjusted to final concentration of 8 mM EGTA and incubated for 1 min at 30 °C, followed by 10 min at 4 °C. Samples for 'inactivation monitoring' were supplemented with 2.5 nM Cy5-labelled RNA at this point. All samples were incubated for a further 30 min at 30 °C. In addition, a control sample without nuclease was prepared.

Samples were deproteinised with stop mix (**Table 7-7**) and the RNA was extracted by PCI before separation on a 10 % denaturing PAGE (1 x TBE, 7 M urea).

7.4 iCLIP Methods

Buffers and sequences for oligonucleotide used in *in vitro* spliceosomal iCLIP are listed in **Table 7-9** and **Table 7-10**, respectively.

Table 7-9: Buffers and reagents for iCLIP experiments. Wash buffers were prepared with Milli Q water (Millipore) and filtered through 0.22 µm. Reaction buffers for enzymes were prepared with RNase free water (Ambion). Reaction mixes based on volumes give the final volume in brackets. *IR-L3 adapter was generated by Martina Hallegger.

buffer	composition
K75	20 mM HEPES pH 7.9, 75 mM KCl, 0.05 % Igepal CA-630, 0.25 mM EDTA, 5 % glycerol
debranching buffer	125 mM KCl, 10 % glycerol, 1 mM DTT, 20 mM HEPES pH 7.5
RNase buffer	50 mM Tris-Cl pH 7.4, 150 mM NaCl, 2 mM EDTA
RNase dilution buffer	10 mM Tris-Cl pH 8.0, 100 mM NaCl, 10 % glycerol
high salt buffer	50 mM Tris-Cl pH 7.4, 1 M NaCl, 1 mM EDTA, 1 % Igepal CA-630, 0.1% SDS, 0.5% sodium deoxycholate
PNK wash buffer	20 mM Tris-Cl pH 7.4, 10 mM $MgCl_2$, 0.2 % Tween-20
PNK irCLIP mix	6 µL 5 x iCLIP PNK buffer, 6 µL PEG400 (Sigma), 1.2 µL T4 PNK, 1.2 µL fastAP, 0.5 µL RNasin (Promega), 15.1 µL H_20 (for 30 µL total)
5 x iCLIP PNK buffer	350 mM Tris-Cl pH 6.5, 50 mM $MgCl_2$, 5 mM DTT
radiolabelling mix	0.6 µL T4 PNK, 1.5 µL γ-^{33}P-ATP, 1.2 µL 10 x reaction buffer (NEB), 8.7 µL H_20 (for 12 µL total)
ligation mix	3 µL 10 x ligation buffer, 0.8 µL DMSO, 2.5 µL T4 RNA ligase 1 (high concentration, NEB), 0.4 µL RNasin, 0.5 µL T4 PNK, 2.5 µL 1 µM pre-adenylated IR-L3 adapter*, 9 µL PEG8,000, 6.3 µL H_20 (for 25 µL total)
10 x ligation buffer	500 mM Tris-Cl pH 7.5, 100 mM $MgCl_2$
adapter removal mix	2 µL buffer 2 (NEB), 0.5 5'deadenylase, 0.5 µL $RecJ_f$ endonuclease, 0.5 µL RNasin, 4 µL PEG400, 12.5 µL H_20 (for 20 µL total)
20 x MOPS buffer	104.6 g MOPS, 60.6 g Tris base, 10.0 g SDS, 3.0 g EDTA in 500 mL total volume (LMB media kitchen)

WB transfer buffer	14.4 g glycine, 3.0 g Tris base, 100 mL methanol for 1 L total volume
PK/SDS buffer	10 mM Tris-Cl pH 7.4, 100 mM NaCl, 1 mM EDTA, 0.2 % SDS
high stringency buffer	15 mM Tris-Cl pH 7.4, 120 mM NaCl, 25 mM KCl, 5 mM EDTA, 1 % Triton X-100 (Sigma), 0.001% SDS, 1 % sodium deoxycholate
1.25 x urea cracking buffer	66.6 mM Tris-Cl, pH 7.4, 8 M urea, 1.33 % SDS
T20IP buffer	50 mM Tris-Cl pH7.4, 150 mM NaCl, 0.5% Tween 20, 0.1 mM EDTA, 1 µL RNasin per 920 µL
crush-soak gel buffer	500mM NaCl, 1mM EDTA, 0.05% SDS

Table 7-10: Oligonucleotides used for sequencing library preparation. The 39 available irCLIP_ddRT primers (## 12-50) vary by the XXXXX unique barcode sequence. WWW is a random mix of A and T, N any nucleotide and iSP18 is an 18-atom hexa-ethylene glycol spacer. The IR-L3 adapter was pre-adenylated and conjugated to IR dye 800 via the internal azide modification by Martina Hallegger.

oligonucleotide	**sequence**
3′ infrared adaptor (IR-L3 adapter)	5′-/5Phos/AGATCGGAAGAGCGGTTCAGAAAAAAAAAAAA/iAzideN/ AAAAAAAAAAAA/3Bio/-3′
irCLIP_ddRT_##	5′-/5Phos/WWW**XXXXX**NNNNAGATCGGAAGAGCGTCGTGAT/iSp18/GGATCC /iSp18/TACTGAACCGC-3′
P3 tall	5′-GCATTCCTGCTGAACCGCTCTTCCGATCT-3′
P3 Solexa	5′-CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGAT CT-3′
P5 Solexa	5′-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGAT CT-3′

7.4.1 Enzymatic tests for individual iCLIP steps

To test digestion by RNase I, 1.7 pmol Cy5-labelled 3xMS2_*UBC4*_AC_short were incubated in 200 µL RNase buffer with different amounts of RNase I for 3 min at 37 °C. The RNase was diluted previously in RNase dilution buffer, adding 2 µL volume per reaction. 50 µL stop mix and deproteinization at 37 °C for 20 min terminated the reaction. 100 µL splice diluent were added before phenol extraction. The RNA was precipitated, dissolved in formamide loading dye and separated on an 8 % denaturing PAGE (1 x TBE, 7 M urea).

To test ligation conditions of the pre-adenylated IR-L3 adapter, 6.75 pmol U5 snRNA A were incubated with 0.5 pmol IR-L3 adapter in 1 x RNA ligase ligation buffer (NEB) with changing amounts of RNA ligase I for 14 hours at different temperatures. The 15 µL reaction was quenched by addition of 10 µL stop mix and incubated for 20 min at 37 °C. The reaction volume was increased by 100 µL splice sample diluent and the nucleic acids purified by phenol extraction. After precipitation, the pellet was dissolved in 10 µL formamide dye, and the nucleic acids separated on an 8 % denaturing PAGE (1 x TBE, 7 M urea). The IRDye 800-labelled species were visualised on an Odyssey CLx (Licor).

To test RNase digestion, dephosphorylation and IR-L3 adapter ligation on the same RNA material, 7.5 pmol Cy5-labelled 3xMS2_*UBC4*_short were either incubated with 0.0375 units RNase I (Ambion) in 50 µL RNase buffer for 3 min at 37 °C or with the same number of units S1 Nuclease (Thermo Scientific) in 50 µL S1 reaction buffer for 15 min at 30 °C. 25 µL stop mix were added to quench the reactions deproteinising the reaction for 20 min at 37 °C. 100 µL splice diluent were added before phenol extraction and precipitation. The pellet was dissolved in 7.5 µL PNK irCLIP mix for 30 min at 37 °C (indicated samples only). After addition of 5 µL stop mix, the RNA was purified again as described above. Pellets were dissolved in 5 µL ligation mix containing the IR-L3 adapter and ligation was performed at 23 °C for 14 h. Samples were purified again and separated on an 8 % denaturing PAGE (1x TBE, 7 M urea).

Debranching was tested under two conditions (a and b). Splicing was performed in 20 µL reactions with 3xMS2_*UBC4*_AC_short pre-mRNA substrate for 30 min at 23 °C. (a) 0.8 µg Dbr1 (Abnova) were added to the reaction and incubated at 30 °C for 20 min. (b) The reaction was deproteinised by incubation with stop mix, followed by PCI extraction and precipitation. The RNA pellet was taken up in K75 buffer supplemented with 2.5 mM $MgCl_2$ and incubated with Dbr1 amounts ranging from 0.04 to 0.8 µg for 20 min at 30 °C. In further tests, K75 buffer was additionally supplemented with different amounts of urea. (a and b) Reactions were added to stop mix, PCI extracted, precipitated and analysed on 10 % denaturing PAGE (1 x TBE, 7 M urea).

7.4.2 *In vitro* spliceosome iCLIP

200 µL spliceosome eluate (compare **7.3.10**, **7.3.11** and **7.3.12** for complex C, C* and P, respectively) were distributed into 40 µL per well of a 96-well plate, placed on an ice-cooled metal block and irradiated with 254 nm light (Stratalinker 2400). UV doses are indicated in the respective chapter. Control samples were kept on ice without irradiation. All subsequent steps were done at 4 °C with cooled buffers. Magnetic beads were shaken during all enzymatic steps.

72 mg urea were added and dissolved for 30 min while shaking. The solution was diluted to 0.8 M urea with K75 buffer, before incubation with anti-FLAG M2 magnetic beads (pre-equilibrated in TPS, 10 µL bead volume for Prp16 and Prp22, 20 µL for SmB and Brr2) for 3 hours. Samples which were subjected to debranching were washed thrice with debranching buffer, before incubation with 3 µL Dbr1 in 50 µL debranching buffer supplemented with 2.5 mM $MgCl_2$ and 3 µL Dbr1 for 20 min at 30 °C.

Beads were washed thrice with RNase buffer. RNase I was diluted in RNase dilution buffer to the optimal concentration (given in the respective chapter), and the digestion was performed in 200 µL RNase buffer for 3 min at 37 °C. Samples were kept at 4 °C for 3 min before washing once with high

salt buffer, and thrice with PNK wash buffer. Beads were transferred to DNA LoBind tubes (Eppendorf) and resuspended in PNK irCLIP mix for dephosphorylation at 37 °C for 40 min. Beads were washed twice with PNK wash buffer. For RNA radiolabelling beads were incubated in 12 µL radiolabelling mix with γ-^{33}P-ATP (Hartmann Analytics or PerkinElmer) for 5 min at 37 °C. Beads were washed once with high salt buffer and thrice with PNK wash buffer. Beads were resuspended in 25 µL ligation mix containing the IR-L3 adapter and ligation was performed for 75 min at 23 °C. After washing once with high salt buffer and twice with PNK buffer, beads were incubated in 20 µL adapter removal reaction mix for 1 hour at 30 °C and 30 min at 37 °C. Early experiments lacked this adapter removal step. After washing twice with PNK buffer, 15 µL 2xNuPAGE™ LDS Sample buffer (Invitrogen) were added to the beads and bound proteins eluted by incubation for <5 min at 70 °C. Samples were loaded on 4-12 % NuPAGE Bis-Tris protein gels (Invitrogen) and run at 180 V with MOPS buffer. Transfer onto nitrocellulose (GE Healthcare Amersham) was performed at 30 V for 90 min with WB transfer buffer. To visualise the radio-signal, the membrane was placed against a phosphor imaging plate for approximately 48 hours, which was then scanned with an Amersham Typhoon (GE Healthcare). The membrane was additionally scanned on an Odyssey CLx (Licor) for detection of the IR-L3 adapter emission signal.

Regions of the target RNA-protein complex were excised from the membrane according to the previously generated images. The same region was extracted for control samples. Nitrocellulose pieces were incubated with 190 µL PK/SDS buffer and 10 µL proteinase K (Roche) for 60 min at 50 °C while shaking. The soluble part was mixed with 200 µL PCI (25:24:1, saturated with 10 mM Tris, pH 8.0, 1mM EDTA, Sigma Aldrich) and the phases were separated by centrifugation in a phase lock gel heavy tube (Eppendorf). The soluble phase was mixed with 800 µL chloroform in the same tube before re-separation by centrifugation. In early experiments, samples were taken before and after PCI extraction, mixed with Ultima Gold liquid scintillation cocktail (PerkinElmer) and measured in a Tri-Carb 2910 liquid scintillation counter (PerkinElmer). Finally, the RNA was precipitated by addition of 0.75 µL GlycoBlue (Ambion), 20 µL 5 M NaCl and 500 µL absolute ethanol at -20 °C for at least 12 hours.

In vitro spliceosome iCLIP experiments for UV and RNase titration were performed similarly. Minimally 40 µL spliceosome eluate were irradiated with different amounts of UV, captured with the relative amount of beads and split afterwards to test different amounts of RNase I. Radiolabelling was performed after RNase digestion, followed by separation on 4-12 % NuPAGE Bis-Tris protein gels. RNase I units were back calculated to represent the equivalent from 200 µL starting volume.

To monitor protein amounts, the nitrocellulose membrane was blocked with milk powder, decorated with anti-FLAG M2-Peroxidase antibody (Sigma Aldrich). Western blots were developed with ECL western blotting detection reagents (GE Healthcare) on film.

7.4.3 Short-iCLIP

Preparation for short-iCLIP samples were performed in the same way up to the IR-L3 adapter ligation step. When this version of the protocol was used, the adapter removal step was not introduced yet. For elution, 80 µL 1.25 x urea cracking buffer was added to the beads and incubated for 3 min at 65 °C while shaking. The supernatant was collected and added to 920 µL T20IP buffer. 50 µL Dynabeads™ Protein G (Invitrogen) were used to preclean the eluate while nutating for 20 min. The supernatant was added to 10 µL anti-FLAG M2 magnetic beads for another 2 hours of capture. Beads were washed once with high salt buffer, and twice with PNK wash buffer. The protein-RNA complexes were directly eluted from the beads by incubation with 190 µL PK/SDS buffer and 10 µL proteinase K for 1 hour at 50 °C. In a second version of the short-iCLIP procedure, precleaning with protein G and the second capture with FLAG beads was omitted. The elution was monitored by scintillation counting of radiolabelled RNA.

7.4.4 Gel-free iCLIP

Preparation for gel-free iCLIP samples followed the same steps as the long protocol until elution from anti-FLAG magnetic beads. To compare on-bead ligation and in-solution ligation, only half of each sample was subjected to the adapter ligation step. Protein-RNA complexes were eluted from anti-FLAG beads by addition of 250 µL RTL buffer (RNEasy Plus Micro Kit supplemented to 40 mM DTT, Qiagen). After removal of the supernatant, the beads were incubated with further 250 µL RTL/DTT buffer and incubated at 37 °C for 2 min while shaking. The combined supernatant was supplemented with 500 µL 70 % EtOH. The RNEasy Plus Micro Kit instructions were followed from step 4 onwards.

For gel-free iCLIP, samples were eluted by adding 20 µL EB (Qiagen) to the RNEasy column, incubated for 2 min at 65 °C while shaking, and released by centrifugation. Samples were kept at -20 °C until transfer to the Francis Crick Institute for further processing. The protein fraction was bound to 2.5 µL SP3 beads (1:1 mix of hydrophilic and hydrophobic beads, GE Healthcare). 200 µL ACN (acetonitrile) were added and the supernatant removed. The beads were washed with another 200 µL ACN. 190 µL PK/SDS buffer and 10 µL proteinase K were added to the beads to digest the protein and release crosslinked RNA during 60 min at 50 °C. The supernatant was cleaned up by PCI extraction, and the RNA was precipitated (as in **7.4.1**). For the second half of the samples, which lacked the IR-L3 adapter, the pellet was resolved in ligation mix and incubated for 75 min at 23 °C.

Agencourt AMPure XP beads (Beckman Coulter) were used for buffer exchange as described in the manufacturer's instructions before the adapter removal reaction. After a second buffer exchange by AMPure XP beads, RNA was eluted in 11 µL H_2O. Reverse transcription reactions for in-solution ligated samples were set up in double volumes later (22 µL compare with **7.4.5**).

7.4.5 Sequencing library preparation

Table 7-11: PCR conditions for cDNA library preparation. RCR conditions are listed with the amounts and composition of the reaction and the run PCR program.

PCR	composition		program
RT	5.5 µL	RNA	65 °C 5 min
	1 µL	primer irCLIP_ddRT_## (1 pmol/µL)	25 °C hold until components
	0.5 µL	dNTP mix (10 mM)	marked with * are added
	2 µL	5 x SSIV buffer (Invitrogen)*	25 °C 5 min
	0.5 µL	0.1 M DTT*	50 °C 5 min
	0.25 µL	RNasin*	55 °C 5 min
	0.25 µL	SuperScript IV (Invitrogen)*	4 °C hold
cDNA elution	1 µL	1 µM P3 tall	95 °C 3 min
	0.75 µL	50 mM $MnCl_2$	95 C -> 60 C
	11 µL	H_2O	temperature ramp @ -0.1 C/sec
			60 °C 5 min
test amplification	1 µL	cDNA	98 C 40 sec
	0.25 µL	primer mix P5 / P3 Solexa, 10 µM each	
	5 µL	Phusion High-Fidelity PCR Master Mix (Thermo Scientific)	15 – 30 cycles:
			98 °C 20 sec
	3.75 µL	H_2O	65 °C 30 sec
			72 °C 45 sec
			72 °C 3 min
			25 °C hold
preparative PCR	4 µL	cDNA	98 °C 40 sec
	1 µL	primer mix P5 Solexa / P3 Solexa, 10 µM each	Optimised number of cycles -2:
	20 µL	Phusion High-Fidelity PCR Master Mix	98 °C 20 sec
	15 µL	H_2O	65 °C 30 sec
			72 °C 45 sec
			72 °C 3 min
			25 °C hold

All sequencing libraries were prepared in the Francis Crick Institute. Used primers and PCR conditions are listed in **Table 7-10** and **Table 7-11**, respectively. The precipitated RNA from *in vitro* spliceosome iCLIP was resolved in H_2O and transcribed into cDNA using distinct irCLIP-RT primers and the SuperScript IV polymerase mix (Invitrogen). The RNA template was digested by addition of 1 µL 1:1 mix of RNase H (NEB) and RNase A (Ambion) for 15 min 37 °C. Up to 3 samples from different experiments were multiplexed at this point. The cDNA was purified by binding to 7.5 µL MyOne C1 SA-dynabeads for 30 min, washed twice with high stringency buffer and thrice with PBS (phos-

phate-buffered saline; Invitrogen). Elution of the cDNA was performed by annealing of the oligonucleotide 'P3 tall' during a temperature ramp from 95 °C to 60 °C, following circularisation by CircLigase II (Epicentre) in CircLigase buffer II for 2 hours at 60 °C.

After further incubation for 3 min at 95 °C, the circularised cDNA product was collected from the beads and precipitated with 0.75 µL GlucoBlue, 0.1 x volume of 3 M sodium acetate and 1.5 x volumes of absolute ethanol at -20 °C. The cDNA pellet was resolved in 10 µL H_2O. Test amplifications (**Table 7-11**) were run on 1 µL of the cDNA and analysed on a 6 % Novex TBE gels (Invitrogen) stained with SYBR green I (Thermo Fisher). This step was repeated with varying cycle numbers until samples were seen at sufficient intensity without over-amplification.

A larger, preparative scale PCR was set up and amplified with the previously determined optimal cycle number minus two, and checked for quality on a 6 % Novex TBE gel. In most cases, the second half of cDNA template was amplified with the same number of PCR cycles. At this point, PCR products of different samples were multiplexed and purified with Agencourt AMPure XP beads, followed by gel extraction from 6 % Novex TBE gel. The amplified DNA was excised from the 145 – 400 nt region. The excised gel piece was crushed and incubated in 500 µL crush-soak gel buffer per gel lane for 2 hours at 65 °C. The liquid portion was cleaned up by centrifugation through two glass prefilters in a Costar SpinX column, followed by precipitation with 1 µL GlycoBlue, 50 µL 3 M sodium acetate pH 5.5 and 1.5 mL absolute ethanol at -20 °C.

7.4.6 Quality control and next-generation sequencing

cDNA pellets were resuspended in 10 – 20 µL H_2O. The concentration was determined with the DeNovix dsDNA High Sensitivity kit according to the manufacturer's instruction. For quality control, the sample was also analysed on the Agilent 2200 TapeStation system using the High Sensitivity D1000 ScreenTape (Agilent).

Libraries were multiplexed with libraries generated from other scientists and analysed by qPCR prior to sequencing at the Illumina HiSeq 2500 platform. qPCR was performed by the scientist responsible for the respective sequencer. Libraries were sequenced with cycle lengths between 100 and 250 nt in single-read mode at the Francis Crick Institute (London), the Gurdon Institute (Cambridge) or the MRC Laboratory of Molecular Biology.

7.4.7 Data Analysis

Data processing was done by Charlotte Capitanchik. A novel Snakemake pipeline (Köster and Rahmann 2012) was written to process all the data in a reproducible manner, with accessory scripts written in R (R Core Team 2017) (the code is attached through a link in the **Appendix**). The first 9 nt of the sequenced read correspond to the barcode, which contains the experimental identifier and

the unique molecular identifier (UMI). Reads were demultiplexed and the 5'- and 3'-adaptor trimmed using iCount demultiplex, which moves the UMI to the fastq header. Reads were then trimmed using Trim Galore! (version 0.4.5, Babraham Bioinformatics) to remove bases from the 3'-end of reads with a quality score of less than phred 20. Reads shorter than 11 nt were removed after the trimming step. To map *in vitro* spliceosome iCLIP data onto the substrate-ome, a novel reference-genome was created for the spliceosomal snRNAs and each transcript used for spliceosome assembly (compare **Table 7-4**). A custom annotation was generated to allow mapping of spliced reads to the canonical junction of each substrate.

The processed reads were then mapped to the substrate-ome using STAR aligner (Dobin et al. 2013), with the following parameters. A maximum of 2 mismatches was allowed and multimappers were removed. The setting `--seedSearchStartLmax 16` was used, meaning that the read is split into more seeds during alignment, resulting in a potentially more sensitive search, which is more suitable for shorter reads. `--alignSJoverhangMin 1000` was used to prevent the generation of spurious splice junctions. Finally, `--alignEndsType EndToEnd` was used to prevent clipping of bases from the 5'-end which could create false crosslink sites. Following mapping to the substrate-ome, reads which remained unmapped or mapped in the reverse orientation were aligned to the *S. cerevisiae* (SacCer3) genome using STAR, allowing for novel splice products, under the assumption that the *S. cerevisiae* annotations may be incomplete.

The UMI was used to quantify numbers of uniquely mapped cDNAs by collapsing cDNAs with the same UMI and which mapped to the same starting position on the substrate-ome or the genome. The -1 position of the read was interpreted as the crosslinked position. Bedgraph-files, coding for the gene-location and number of crosslink-events, were visualised in Integrative Genomics Viewer (IGV) (Robinson et al. 2011), traces exported and figures prepared in Illustrator (Adobe).

To determine the mutation rate introduced at the 5'SS and the brA during RT, the bam-files of experiments conducted with 3xMS2_*UBC4*_(AC/WT)_long were loaded into IGV. The control experiment (no_FLAGged_protein) was excluded. At the 5'SS the position **x**|GU (wherein | indicates the splice site) was analysed, and at the brA the read accuracy for the branch point adenosine. The base identity (in percent) was extracted from IGV directly. Determination of mean and median were determined in Origin 8.0 (OriginLab, Northampton, MA) and illustrated in a box plot.

To assess if the amount of truncation events between the downstream of the brA is pronounced in UV-irradiated samples, the sum of truncation events of position 234 to 279 nt was divided by the

number of truncation events at position 162 corresponding to the 5'SS. The respective numbers were extracted from the bedgraph-files using Python (Python Software Foundation) and plotted with Origin 8.0.

Read coverage was calculated from bam-files using the depth command of SAMtools (version 0.1.19) (Li et al. 2009). Coverage ratios were subsequently determined by dividing the coverage of the regions downstream of brA (nucleotide 232 in 3xMS2_*UBC4*_short) by the coverage of the region upstream of brA and plotted in R (R Core Team 2017).

Acknowledgements

The time of my predoctoral research was both an opportunity to improve as a scientist and grow personally. I want to thank all the people without whom this journey would not have been possible.

Firstly, I want to thank my supervisor Dr Kiyoshi Nagai for his support and guidance throughout the whole time. Kiyoshi, thank you for your trust and the faith which you put into '*in vitro* spliceosome iCLIP', a project that still had to prove its worth. Secondly, I am grateful to Prof. Jernej Ule, who generously hosted me in the Crick Institute, making me feel like a full lab member. Thank you for the many discussions and your constant and infectious enthusiasm for my project and science in general. In addition, I want to thank Dr Andrew Newman, who contributed enormously to this project with both scientific knowledge and technical help. Thank you, Andy, also for proofreading my whole thesis. At the same time, I am also grateful to have worked together with Martina Hallegger, who taught me everything about iCLIP and with Charlotte Capitanchik, who turned '*in vitro* spliceosome iCLIP' into a real team effort. It was enriching to working with you and I look forward to continuing our collaboration. I want to thank Philipp Holliger and Ben Porebski, who offered to sequence my samples at the LMB, which really speeded up my project. Additionally, I am very thankful for having wonderful lab mates in both the Nagai and the Ule lab, as well as the wider LMB. To name a few, I am grateful for the help and support of:

Chris Oubridge, with who I worked on the Snu114 crystallography project. Chris Norman, for generating some yeast strains, testing the Dbr1 enzyme and helping whenever time was short. Ben Porebski, who taught me in Python scripting. Clemens Plaschka, who never got tired discussing tiniest details with me when I needed a second opinion. Migle Kisonaite and Suyang Zhang, for sharing both the moments of disappointment when little progress was made and the moments of excitement about meaningful results.

I also want to thank my second supervisors, Dr Simon Bullock and Prof. Ben Luisi for their advice and support. Furthermore, I want to thank the Boehringer Ingelheim Fonds and the MRC Laboratory of Molecular Biology for funding. I also want to thank my friends and housemates, who challenged me to explain my research to a non-expert audience, helping me step back and see scientific problems in a new light. Finally, I am grateful to have my family and my boyfriend Thomas, who constantly supported me and encouraged my future plans and dreams.

Bibliography

Abovich, N., X. C. Liao, and M. Rosbash. 1994. 'The yeast MUD2 protein: an interaction with PRP11 defines a bridge between commitment complexes and U2 snRNP addition', *Genes Dev*, 8: 843-54.

Aronova, A., D. Bacikova, L. B. Crotti, D. S. Horowitz, and B. Schwer. 2007. 'Functional interactions between Prp8, Prp18, Slu7, and U5 snRNA during the second step of pre-mRNA splicing', *RNA (New York, N.Y.)*, 13: 1437-44.

Bai, R., R. Wan, C. Yan, J. Lei, and Y. Shi. 2018. 'Structures of the fully assembled Saccharomyces cerevisiae spliceosome before activation', *Science*.

Bai, R., C. Yan, R. Wan, J. Lei, and Y. Shi. 2017. 'Structure of the Post-catalytic Spliceosome from Saccharomyces cerevisiae', *Cell*, 171: 1589-98.e8.

Bartels, C., C. Klatt, R. Luhrmann, and P. Fabrizio. 2002. 'The ribosomal translocase homologue Snu114p is involved in unwinding U4/U6 RNA during activation of the spliceosome', *EMBO Rep*, 3: 875-80.

Bartels, C., H. Urlaub, R. Luhrmann, and P. Fabrizio. 2003. 'Mutagenesis suggests several roles of Snu114p in pre-mRNA splicing', *J Biol Chem*, 278: 28324-34.

Berglund, J. A., K. Chua, N. Abovich, R. Reed, and M. Rosbash. 1997. 'The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC', *Cell*, 89: 781-7.

Bertram, K., D. E. Agafonov, O. Dybkov, D. Haselbach, M. N. Leelaram, C. L. Will, H. Urlaub, B. Kastner, R. Luhrmann, and H. Stark. 2017. 'Cryo-EM Structure of a Pre-catalytic Human Spliceosome Primed for Activation', *Cell*, 170: 701-13.e11.

Bessonov, S., M. Anokhina, C. L. Will, H. Urlaub, and R. Luhrmann. 2008. 'Isolation of an active step I spliceosome and composition of its RNP core', *Nature*, 452: 846-50.

Bindereif, A., and M. R. Green. 1987. 'An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly', *Embo J*, 6: 2415-24.

Bishop, A., O. Buzko, S. Heyeck-Dumas, I. Jung, B. Kraybill, Y. Liu, K. Shah, S. Ulrich, L. Witucki, F. Yang, C. Zhang, and K. M. Shokat. 2000. 'Unnatural ligands for engineered proteins: new tools for chemical genetics', *Annu Rev Biophys Biomol Struct*, 29: 577-606.

Boesler, C., N. Rigo, D. E. Agafonov, B. Kastner, H. Urlaub, C. L. Will, and R. Luhrmann. 2015. 'Stable tri-snRNP integration is accompanied by a major structural rearrangement of the spliceosome that is dependent on Prp8 interaction with the 5' splice site', *RNA (New York, N.Y.)*, 21: 1993-2005.

Boesler, C., N. Rigo, M. M. Anokhina, M. J. Tauchert, D. E. Agafonov, B. Kastner, H. Urlaub, R. Ficner, C. L. Will, and R. Luhrmann. 2016. 'A spliceosome intermediate with loosely associated tri-snRNP accumulates in the absence of Prp28 ATPase activity', *Nat Commun*, 7: 11997.

Boon, K. L., C. M. Norman, R. J. Grainger, A. J. Newman, and J. D. Beggs. 2006. 'Prp8p dissection reveals domain structure and protein interaction sites', *RNA (New York, N.Y.)*, 12: 198-205.

Brenner, T. J., and C. Guthrie. 2005. 'Genetic analysis reveals a role for the C terminus of the Saccharomyces cerevisiae GTPase Snu114 during spliceosome activation', *Genetics*, 170: 1063-80.

Briese, Michael, Nejc Haberman, Christopher Sibley, Anob Chakrabarti, Zhen Wang, Julian Konig, David Perera, Vihandha Wickramasinghe, Ashok Venkitaraman, Nicholas Luscombe, Christopher Smith, Tomaz Curk, and Jernej Ule. 2018. 'A systems view of spliceosomal assembly and branchpoints with iCLIP', *bioRxiv*: 353599.

Burgess, S. M., and C. Guthrie. 1993. 'A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates', *Cell*, 73: 1377-91.

Burke, J. E., A. D. Longhurst, D. Merkurjev, J. Sales-Lee, B. Rao, J. J. Moresco, J. R. Yates, 3rd, J. J. Li, and H. D. Madhani. 2018. 'Spliceosome Profiling Visualizes Operations of a Dynamic RNP at Nucleotide Resolution', *Cell*, 173: 1014-30.e17.

Buttner, K., S. Nehring, and K. P. Hopfner. 2007. 'Structural basis for DNA duplex separation by a superfamily-2 helicase', *Nat Struct Mol Biol*, 14: 647-52.

Caruthers, J. M., and D. B. McKay. 2002. 'Helicase structure and mechanism', *Curr Opin Struct Biol*, 12: 123-33.

Chakrabarti, Anob M., Nejc Haberman, Arne Praznik, Nicholas M. Luscombe, and Jernej Ule. 2017. 'Data Science Issues in Understanding Protein-RNA Interactions', *bioRxiv*.

Chan, S. P., and S. C. Cheng. 2005. 'The Prp19-associated complex is required for specifying interactions of U5 and U6 with pre-mRNA during spliceosome activation', *J Biol Chem*, 280: 31190-9.

Chan, S. P., D. I. Kao, W. Y. Tsai, and S. C. Cheng. 2003. 'The Prp19p-associated complex in spliceosome activation', *Science*, 302: 279-82.

Chapman, Karen B., and Jef D. Boeke. 1991. 'Isolation and characterization of the gene encoding yeast debranching enzyme', *Cell*, 65: 483-92.

Chen, H. C., C. K. Tseng, R. T. Tsai, C. S. Chung, and S. C. Cheng. 2013. 'Link of NTR-mediated spliceosome disassembly with DEAH-box ATPases Prp2, Prp16, and Prp22', *Mol Cell Biol*, 33: 514-25.

Chen, W., J. Moore, H. Ozadam, H. P. Shulha, N. Rhind, Z. Weng, and M. J. Moore. 2018. 'Transcriptome-wide Interrogation of the Functional Intronome by Spliceosome Profiling', *Cell*, 173: 1031-44.e13.

Cheng, S. C., and J. Abelson. 1986. 'Fractionation and characterization of a yeast mRNA splicing extract', *Proc Natl Acad Sci U S A*, 83: 2387-91.

Cherfils, J., and M. Zeghouf. 2013. 'Regulation of small GTPases by GEFs, GAPs, and GDIs', *Physiol Rev*, 93: 269-309.

Chiu, Y. F., Y. C. Liu, T. W. Chiang, T. C. Yeh, C. K. Tseng, N. Y. Wu, and S. C. Cheng. 2009. 'Cwc25 is a novel splicing factor required after Prp2 and Yju2 to facilitate the first catalytic reaction', *Mol Cell Biol*, 29: 5671-8.

Company, M., J. Arenas, and J. Abelson. 1991. 'Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes', *Nature*, 349: 487-93.

Cordin, O., J. Banroques, N. K. Tanner, and P. Linder. 2006. 'The DEAD-box protein family of RNA helicases', *Gene*, 367: 17-37.

Cordin, O., and J. D. Beggs. 2013. 'RNA helicases in splicing', *RNA Biol*, 10: 83-95.

DeHaven, A. C., I. S. Norden, and A. A. Hoskins. 2016. 'Lights, camera, action! Capturing the spliceosome and pre-mRNA splicing with single-molecule fluorescence microscopy', *Wiley Interdiscip Rev RNA*, 7: 683-701.

Dix, I., C. S. Russell, R. T. O'Keefe, A. J. Newman, and J. D. Beggs. 1998. 'Protein-RNA interactions in the U5 snRNP of Saccharomyces cerevisiae', *RNA (New York, N.Y.)*, 4: 1675-86.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics*, 29: 15-21.

Domdey, H., B. Apostol, R. J. Lin, A. Newman, E. Brody, and J. Abelson. 1984. 'Lariat structures are in vivo intermediates in yeast pre-mRNA splicing', *Cell*, 39: 611-21.

Fabrizio, P., J. Dannenberg, P. Dube, B. Kastner, H. Stark, H. Urlaub, and R. Luhrmann. 2009. 'The evolutionarily conserved core design of the catalytic activation step of the yeast spliceosome', *Mol Cell*, 36: 593-608.

Fabrizio, P., B. Laggerbauer, J. Lauber, W. S. Lane, and R. Luhrmann. 1997. 'An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2', *Embo J*, 16: 4092-106.

Fairman-Williams, Margaret E., Ulf-Peter Guenther, and Eckhard Jankowsky. 2010. 'SF1 and SF2 helicases: family matters', *Current Opinion in Structural Biology*, 20: 313-24.

Fica, S. M., M. A. Mefford, J. A. Piccirilli, and J. P. Staley. 2014. 'Evidence for a group II intron-like catalytic triplex in the spliceosome', *Nat Struct Mol Biol*, 21: 464-71.

Fica, S. M., C. Oubridge, W. P. Galej, M. E. Wilkinson, X. C. Bai, A. J. Newman, and K. Nagai. 2017. 'Structure of a spliceosome remodelled for exon ligation', *Nature*, 542: 377-80.

Fica, S. M., N. Tuttle, T. Novak, N. S. Li, J. Lu, P. Koodathingal, Q. Dai, J. P. Staley, and J. A. Piccirilli. 2013. 'RNA catalyses nuclear pre-mRNA splicing', *Nature*, 503: 229-34.

Fica, Sebastian M., and Kiyoshi Nagai. 2017. 'Cryo-electron microscopy snapshots of the spliceosome: structural insights into a dynamic ribonucleoprotein machine', *Nat Struct Mol Biol*, 24: 791-99.

Fischer, C. J., and T. M. Lohman. 2004. 'ATP-dependent translocation of proteins along single-stranded DNA: models and methods of analysis of pre-steady state kinetics', *J Mol Biol*, 344: 1265-86.

Fleckner, J., M. Zhang, J. Valcarcel, and M. R. Green. 1997. 'U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction', *Genes Dev*, 11: 1864-72.

Fourmann, J. B., O. Dybkov, D. E. Agafonov, M. J. Tauchert, H. Urlaub, R. Ficner, P. Fabrizio, and R. Luhrmann. 2016. 'The target of the DEAH-box NTP triphosphatase Prp43 in Saccharomyces cerevisiae spliceosomes is the U2 snRNP-intron interaction', *Elife*, 5.

Fourmann, J. B., J. Schmitzova, H. Christian, H. Urlaub, R. Ficner, K. L. Boon, P. Fabrizio, and R. Luhrmann. 2013. 'Dissection of the factor requirements for spliceosome disassembly and the elucidation of its dissociation products using a purified splicing system', *Genes Dev*, 27: 413-28.

Frank, D., and C. Guthrie. 1992. 'An essential splicing factor, SLU7, mediates 3' splice site choice in yeast', *Genes Dev*, 6: 2112-24.

Frank, J. 2017. 'Time-resolved cryo-electron microscopy: Recent progress', *J Struct Biol*, 200: 303-06.

Frazer, L. N., S. C. Lovell, and R. T. O'Keefe. 2009. 'Analysis of synthetic lethality reveals genetic interactions between the GTPase Snu114p and snRNAs in the catalytic core of the Saccharomyces cerevisiae spliceosome', *Genetics*, 183: 497-515-1si-4si.

Furger, A., J. M. O'Sullivan, A. Binnie, B. A. Lee, and N. J. Proudfoot. 2002. 'Promoter proximal splice sites enhance transcription', *Genes Dev*, 16: 2792-9.

Galej, W. P., C. Oubridge, A. J. Newman, and K. Nagai. 2013. 'Crystal structure of Prp8 reveals active site cavity of the spliceosome', *Nature*, 493: 638-43.

Galej, W. P., M. E. Wilkinson, S. M. Fica, C. Oubridge, A. J. Newman, and K. Nagai. 2016. 'Cryo-EM structure of the spliceosome immediately after branching', *Nature*, 537: 197-201.

Gilman, B., P. Tijerina, and R. Russell. 2017. 'Distinct RNA-unwinding mechanisms of DEAD-box and DEAH-box RNA helicase proteins in remodeling structured RNAs and RNPs', *Biochem Soc Trans*, 45: 1313-21.

Grainger, R. J., J. D. Barrass, A. Jacquier, J. C. Rain, and J. D. Beggs. 2009. 'Physical and genetic interactions of yeast Cwc21p, an ortholog of human SRm300/SRRM2, suggest a role at the catalytic center of the spliceosome', *RNA (New York, N.Y.)*, 15: 2161-73.

Guthrie, C., and B. Patterson. 1988. 'Spliceosomal snRNAs', *Annu Rev Genet*, 22: 387-419.

Haberman, N., I. Huppertz, J. Attig, J. Konig, Z. Wang, C. Hauer, M. W. Hentze, A. E. Kulozik, H. Le Hir, T. Curk, C. R. Sibley, K. Zarnack, and J. Ule. 2017. 'Insights into the design and interpretation of iCLIP experiments', *Genome Biol*, 18: 7.

Hahn, D., G. Kudla, D. Tollervey, and J. D. Beggs. 2012. 'Brr2p-mediated conformational rearrangements in the spliceosome during activation and substrate repositioning', *Genes Dev*, 26: 2408-21.

Hall, M. C., and S. W. Matson. 1999. 'Helicase motifs: the engine that powers DNA unwinding', *Mol Microbiol*, 34: 867-77.

Hallegger, M., M. Llorian, and C. W. Smith. 2010. 'Alternative splicing: global insights', *Febs j*, 277: 856-66.

Hang, J., R. Wan, C. Yan, and Y. Shi. 2015. 'Structural basis of pre-mRNA splicing', *Science*, 349: 1191-8.

He, Y., G. R. Andersen, and K. H. Nielsen. 2010. 'Structural basis for the function of DEAH helicases', *EMBO Rep*, 11: 180-6.

He, Y., J. P. Staley, G. R. Andersen, and K. H. Nielsen. 2017. 'Structure of the DEAH/RHA ATPase Prp43p bound to RNA implicates a pair of hairpins and motif Va in translocation along RNA', *RNA (New York, N.Y.)*, 23: 1110-24.

Henning, L. M., K. F. Santos, J. Sticht, S. Jehle, C. T. Lee, M. Wittwer, H. Urlaub, U. Stelzl, M. C. Wahl, and C. Freund. 2017. 'A new role for FBP21 as regulator of Brr2 helicase activity', *Nucleic Acids Res*, 45: 7922-37.

Hoskins, A. A., J. Gelles, and M. J. Moore. 2011. 'New insights into the spliceosome by single molecule fluorescence microscopy', *Curr Opin Chem Biol*, 15: 864-70.

Hoskins, A. A., M. L. Rodgers, L. J. Friedman, J. Gelles, and M. J. Moore. 2016. 'Single molecule analysis reveals reversible and irreversible steps during spliceosome activation', *Elife*, 5.

Huppertz, I., J. Attig, A. D'Ambrogio, L. E. Easton, C. R. Sibley, Y. Sugimoto, M. Tajnik, J. Konig, and J. Ule. 2014. 'iCLIP: protein-RNA interactions at nucleotide resolution', *Methods*, 65: 274-87.

Jacewicz, A., B. Schwer, P. Smith, and S. Shuman. 2014. 'Crystal structure, mutational analysis and RNA-dependent ATPase activity of the yeast DEAD-box pre-mRNA splicing factor Prp28', *Nucleic Acids Res*, 42: 12885-98.

James, S. A., W. Turner, and B. Schwer. 2002. 'How Slu7 and Prp18 cooperate in the second step of yeast pre-mRNA splicing', *RNA (New York, N.Y.)*, 8: 1068-77.

Jankowsky, E. 2011. 'RNA helicases at work: binding and rearranging', *Trends Biochem Sci*, 36: 19-29.

Jankowsky, E., C. H. Gross, S. Shuman, and A. M. Pyle. 2001. 'Active disruption of an RNA-protein interaction by a DExH/D RNA helicase', *Science*, 291: 121-5.

Jones, M. H., and C. Guthrie. 1990. 'Unexpected flexibility in an evolutionarily conserved protein-RNA interaction: genetic analysis of the Sm binding site', *Embo J*, 9: 2555-61.

Kim, D. H., and J. J. Rossi. 1999. 'The first ATPase domain of the yeast 246-kDa protein is required for in vivo unwinding of the U4/U6 duplex', *RNA (New York, N.Y.)*, 5: 959-71.

Kim, S. H., and R. J. Lin. 1996. 'Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing', *Mol Cell Biol*, 16: 6810-9.

Kim, S. H., J. Smith, A. Claude, and R. J. Lin. 1992. 'The purified yeast pre-mRNA splicing factor PRP2 is an RNA-dependent NTPase', *Embo J*, 11: 2319-26.

Kistler, A. L., and C. Guthrie. 2001. 'Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for Sub2, an essential spliceosomal ATPase', *Genes Dev*, 15: 42-9.

Konig, J., K. Zarnack, G. Rot, T. Curk, M. Kayikci, B. Zupan, D. J. Turner, N. M. Luscombe, and J. Ule. 2010. 'iCLIP reveals the function of hnRNP particles in splicing at individual nucleotide resolution', *Nat Struct Mol Biol*, 17: 909-15.

Koodathingal, P., T. Novak, J. A. Piccirilli, and J. P. Staley. 2010. 'The DEAH box ATPases Prp16 and Prp43 cooperate to proofread 5' splice site cleavage during pre-mRNA splicing', *Mol Cell*, 39: 385-95.

Kosowski, Tomasz R., Heather R. Keys, Tiffani K. Quan, and Stephanie W. Ruby. 2009. 'DExD/H-box Prp5 protein is in the spliceosome during most of the splicing cycle', *RNA (New York, N.Y.)*, 15: 1345-62.

Köster, Johannes, and Sven Rahmann. 2012. 'Snakemake—a scalable bioinformatics workflow engine', *Bioinformatics*, 28: 2520-22.

Krishnan, R., M. R. Blanco, M. L. Kahlscheuer, J. Abelson, C. Guthrie, and N. G. Walter. 2013. 'Biased Brownian ratcheting leads to pre-mRNA remodeling and capture prior to first-step splicing', *Nat Struct Mol Biol*, 20: 1450-7.

Kunkel, T. A. 1985. 'Rapid and efficient site-specific mutagenesis without phenotypic selection', *Proc Natl Acad Sci U S A.*, 82: 488-92.

Laggerbauer, B., T. Achsel, and R. Luhrmann. 1998. 'The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro', *Proc Natl Acad Sci U S A*, 95: 4188-92.

Lardelli, R. M., J. X. Thompson, J. R. Yates, 3rd, and S. W. Stevens. 2010. 'Release of SF3 from the intron branchpoint activates the first step of pre-mRNA splicing', *RNA (New York, N.Y.)*, 16: 516-28.

Larson, J. D., and A. A. Hoskins. 2017. 'Dynamics and consequences of spliceosome E complex formation', *Elife*, 6.

Lauber, J., P. Fabrizio, S. Teigelkamp, W. S. Lane, E. Hartmann, and R. Luhrmann. 1996. 'The HeLa 200 kDa U5 snRNP-specific protein and its homologue in Saccharomyces cerevisiae are members of the DEXH-box protein family of putative RNA helicases', *Embo J*, 15: 4001-15.

Le Hir, H., D. Gatfield, I. C. Braun, D. Forler, and E. Izaurralde. 2001. 'The protein Mago provides a link between splicing and mRNA localization', *EMBO Rep*, 2: 1119-24.

Lerner, M. R., and J. A. Steitz. 1979. 'Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus', *Proc Natl Acad Sci U S A*, 76: 5495-9.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics*, 25: 2078-9.

Liang, W. W., and S. C. Cheng. 2015. 'A novel mechanism for Prp5 function in prespliceosome formation and proofreading the branch site sequence', *Genes Dev*, 29: 81-93.

Lin, J., M. G. Gagnon, D. Bulkley, and T. A. Steitz. 2015. 'Conformational changes of elongation factor G on the ribosome during tRNA translocation', *Cell*, 160: 219-27.

Lin, R. J., A. J. Newman, S. C. Cheng, and J. Abelson. 1985. 'Yeast mRNA splicing in vitro', *J Biol Chem*, 260: 14780-92.

Linder, Patrick. 2006. 'Dead-box proteins: a family affair--active and passive players in RNP-remodeling', *Nucleic Acids Research*, 34: 4168-80.

Liu, S., X. Li, L. Zhang, J. Jiang, R. C. Hill, Y. Cui, K. C. Hansen, Z. H. Zhou, and R. Zhao. 2017. 'Structure of the yeast spliceosomal postcatalytic P complex', *Science*, 358: 1278-83.

Lucius, A. L., N. K. Maluf, C. J. Fischer, and T. M. Lohman. 2003. 'General methods for analysis of sequential "n-step" kinetic mechanisms: application to single turnover kinetics of helicase-catalyzed DNA unwinding', *Biophys J*, 85: 2224-39.

Luukkonen, B. G., and B. Seraphin. 1997. 'The role of branchpoint-3' splice site spacing and interaction between intron terminal nucleotides in 3' splice site selection in Saccharomyces cerevisiae', *Embo J*, 16: 779-92.

Madhani, H. D., and C. Guthrie. 1992. 'A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome', *Cell*, 71: 803-17.

Makarov, E. M., O. V. Makarova, H. Urlaub, M. Gentzel, C. L. Will, M. Wilm, and R. Luhrmann. 2002. 'Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome', *Science*, 298: 2205-8.

Mayas, R. M., H. Maita, D. R. Semlow, and J. P. Staley. 2010. 'Spliceosome discards intermediates via the DEAH box ATPase Prp43p', *Proc Natl Acad Sci U S A*, 107: 10020-5.

Mayas, R. M., H. Maita, and J. P. Staley. 2006. 'Exon ligation is proofread by the DExD/H-box ATPase Prp22p', *Nat Struct Mol Biol*, 13: 482-90.

McPheeters, D. S., and P. Muhlenkamp. 2003. 'Spatial organization of protein-RNA interactions in the branch site-3' splice site region during pre-mRNA splicing in yeast', *Mol Cell Biol*, 23: 4174-86.

McPheeters, D. S., B. Schwer, and P. Muhlenkamp. 2000. 'Interaction of the yeast DExH-box RNA helicase Prp22p with the 3' splice site during the second step of nuclear pre-mRNA splicing', *Nucleic Acids Res*, 28: 1313-21.

Mefford, M. A., and J. P. Staley. 2009. 'Evidence that U2/U6 helix I promotes both catalytic steps of pre-mRNA splicing and rearranges in between these steps', *RNA (New York, N.Y.)*, 15: 1386-97.

Moore, M. J., and N. J. Proudfoot. 2009. 'Pre-mRNA processing reaches back to transcription and ahead to translation', *Cell*, 136: 688-700.

Moore, M. J., and P. A. Sharp. 1992. 'Site-specific modification of pre-mRNA: the 2'-hydroxyl groups at the splice sites', *Science*, 256: 992-7.

Mozaffari-Jovin, S., K. F. Santos, H. H. Hsiao, C. L. Will, H. Urlaub, M. C. Wahl, and R. Luhrmann. 2012. 'The Prp8 RNase H-like domain inhibits Brr2-mediated U4/U6 snRNA unwinding by blocking Brr2 loading onto the U4 snRNA', *Genes Dev*, 26: 2422-34.

Mozaffari-Jovin, S., T. Wandersleben, K. F. Santos, C. L. Will, R. Luhrmann, and M. C. Wahl. 2013. 'Inhibition of RNA helicase Brr2 by the C-terminal tail of the spliceosomal protein Prp8', *Science*, 341: 80-4.

Nancollis, V., J. P. Ruckshanthi, L. N. Frazer, and R. T. O'Keefe. 2013. 'The U5 snRNA internal loop 1 is a platform for Brr2, Snu114 and Prp8 protein binding during U5 snRNP assembly', *J Cell Biochem*, 114: 2770-84.

Newman, A. J., R. J. Lin, S. C. Cheng, and J. Abelson. 1985. 'Molecular consequences of specific intron mutations on yeast mRNA splicing in vivo and in vitro', *Cell*, 42: 335-44.

Newman, A. J., and C. Norman. 1992. 'U5 snRNA interacts with exon sequences at 5' and 3' splice sites', *Cell*, 68: 743-54.

Nguyen, T. H., W. P. Galej, X. C. Bai, C. Oubridge, A. J. Newman, S. H. Scheres, and K. Nagai. 2016. 'Cryo-EM structure of the yeast U4/U6.U5 tri-snRNP at 3.7 Å resolution', *Nature*, 1.

Nguyen, T. H., W. P. Galej, X. C. Bai, C. G. Savva, A. J. Newman, S. H. Scheres, and K. Nagai. 2015. 'The architecture of the spliceosomal U4/U6.U5 tri-snRNP', *Nature*, 523: 47-52.

Nielsen, K. H., and J. P. Staley. 2012. 'Spliceosome activation: U4 is the path, stem I is the goal, and Prp8 is the keeper. Let's cheer for the ATPase Brr2!', *Genes Dev*, 26: 2461-7.

O'Day, C. L., G. Dalbadie-McFarland, and J. Abelson. 1996. 'The Saccharomyces cerevisiae Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA', *J Biol Chem*, 271: 33261-7.

O'Keefe, R. T., C. Norman, and A. J. Newman. 1996. 'The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast', *Cell*, 86: 679-89.

Ohrt, T., P. Odenwalder, J. Dannenberg, M. Prior, Z. Warkocki, J. Schmitzova, R. Karaduman, I. Gregor, J. Enderlein, P. Fabrizio, and R. Luhrmann. 2013. 'Molecular dissection of step 2 catalysis of yeast pre-mRNA splicing investigated in a purified system', *RNA (New York, N.Y.)*, 19: 902-15.

Ooi, S. L., C. Dann, 3rd, K. Nam, D. J. Leahy, M. J. Damha, and J. D. Boeke. 2001. 'RNA lariat debranching enzyme', *Methods Enzymol*, 342: 233-48.

Pang, P. S., E. Jankowsky, P. J. Planet, and A. M. Pyle. 2002. 'The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding', *Embo J*, 21: 1168-76.

Pawlicki, J. M., and J. A. Steitz. 2010. 'Nuclear networking fashions pre-messenger RNA and primary microRNA transcripts for function', *Trends Cell Biol*, 20: 52-61.

Perriman, R., and M. Ares, Jr. 2010. 'Invariant U2 snRNA nucleotides form a stem loop to recognize the intron early in splicing', *Mol Cell*, 38: 416-27.

Plaschka, C., P. C. Lin, and K. Nagai. 2017. 'Structure of a pre-catalytic spliceosome', *Nature*, 546: 617-21.

Plumpton, M., M. McGarvey, and J. D. Beggs. 1994. 'A dominant negative mutation in the conserved RNA helicase motif 'SAT' causes splicing factor PRP2 to stall in spliceosomes', *Embo J*, 13: 879-87.

Polayes, D. A., T. D. Parks, S. A. Johnston, and W. G. Dougherty. 1998. 'Application of TEV Protease in Protein Production', *Methods Mol Med*, 13: 169-83.

Price, S. R., N. Ito, C. Oubridge, J. M. Avis, and K. Nagai. 1995. 'Crystallization of RNA-protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies', *J Mol Biol*, 249: 398-408.

Pyle, A. M. 2008. 'Translocation and unwinding mechanisms of RNA and DNA helicases', *Annu Rev Biophys*, 37: 317-36.

Query, C. C., and M. M. Konarska. 2004. 'Suppression of multiple substrate mutations by spliceosomal prp8 alleles suggests functional correlations with ribosomal ambiguity mutants', *Mol Cell*, 14: 343-54.

R Core Team. 2017. "R: A Language and Environment for Statistical Computing." In.

Raghunathan, P. L., and C. Guthrie. 1998. 'RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2', *Curr Biol*, 8: 847-55.

Reed, R. 1989. 'The organization of 3' splice-site sequences in mammalian introns', *Genes Dev*, 3: 2113-23.

Roberts, G. C., and C. W. Smith. 2002. 'Alternative splicing: combinatorial output from the genome', *Curr Opin Chem Biol*, 6: 375-83.

Robinson, James T., Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, and Jill P. Mesirov. 2011. 'Integrative genomics viewer', *Nature Biotechnology*, 29: 24.

Santos, K. F., S. M. Jovin, G. Weber, V. Pena, R. Luhrmann, and M. C. Wahl. 2012. 'Structural basis for functional cooperation between tandem helicase cassettes in Brr2-mediated remodeling of the spliceosome', *Proc Natl Acad Sci U S A*, 109: 17418-23.

Schneider, Cornelius, Dmitry E. Agafonov, Jana Schmitzová, Klaus Hartmuth, Patrizia Fabrizio, and Reinhard Lührmann. 2015. 'Dynamic Contacts of U2, RES, Cwc25, Prp8 and Prp45 Proteins with the Pre-mRNA Branch-Site and 3' Splice Site during Catalytic Activation and Step 1 Catalysis in Yeast Spliceosomes', *PLoS Genet*, 11: e1005539.

Schneider, Susanne, Hans-Rudolf Hotz, and Beate Schwer. 2002. 'Characterization of Dominant-negative Mutants of the DEAH-box Splicing Factors Prp22 and Prp16', *Journal of Biological Chemistry*, 277: 15452-58.

Schwer, B. 2008. 'A conformational rearrangement in the spliceosome sets the stage for Prp22-dependent mRNA release', *Mol Cell*, 30: 743-54.

Schwer, B., and C. H. Gross. 1998. 'Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing', *Embo J*, 17: 2086-94.

Schwer, B., and C. Guthrie. 1991. 'PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome', *Nature*, 349: 494-9.

———. 1992a. 'A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis', *Embo J*, 11: 5033-9.

———. 1992b. 'A dominant negative mutation in a spliceosomal ATPase affects ATP hydrolysis but not binding to the spliceosome', *Mol Cell Biol*, 12: 3540-7.

Seipelt, Rebecca L., Binhai Zheng, Agatha Asuru, and Brian C. Rymond. 1999. 'U1 snRNA is cleaved by RNase III and processed through an Sm site-dependent pathway', *Nucleic Acids Research*, 27: 587-95.

Semlow, D. R., M. R. Blanco, N. G. Walter, and J. P. Staley. 2016. 'Spliceosomal DEAH-Box ATPases Remodel Pre-mRNA to Activate Alternative Splice Sites', *Cell*, 164: 985-98.

Semlow, D. R., and J. P. Staley. 2012. 'Staying on message: ensuring fidelity in pre-mRNA splicing', *Trends Biochem Sci*, 37: 263-73.

Sharp, P. A. 1987. 'Splicing of messenger RNA precursors', *Science*, 235: 766-71.

Sheth, N., X. Roca, M. L. Hastings, T. Roeder, A. R. Krainer, and R. Sachidanandam. 2006. 'Comprehensive splice-site analysis using comparative genomics', *Nucleic Acids Res*, 34: 3955-67.

Siliciano, P. G., and C. Guthrie. 1988. '5' splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements', *Genes Dev*, 2: 1258-67.

Singleton, M. R., M. S. Dillingham, and D. B. Wigley. 2007. 'Structure and mechanism of helicases and nucleic acid translocases', *Annu Rev Biochem*, 76: 23-50.

Singleton, M. R., and D. B. Wigley. 2002. 'Modularity and specialization in superfamily 1 and 2 helicases', *J Bacteriol*, 184: 1819-26.

Small, E. C., S. R. Leggett, A. A. Winans, and J. P. Staley. 2006. 'The EF-G-like GTPase Snu114p regulates spliceosome dynamics mediated by Brr2p, a DExD/H box ATPase', *Mol Cell*, 23: 389-99.

Sontheimer, E. J., and J. A. Steitz. 1993. 'The U5 and U6 small nuclear RNAs as active site components of the spliceosome', *Science*, 262: 1989-96.

Spingola, M., L. Grate, D. Haussler, and M. Ares, Jr. 1999. 'Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae', *RNA (New York, N.Y.)*, 5: 221-34.

Staley, J. P., and C. Guthrie. 1998. 'Mechanical devices of the spliceosome: motors, clocks, springs, and things', *Cell*, 92: 315-26.

———. 1999. 'An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p', *Mol Cell*, 3: 55-64.

Sugimoto, Y., J. Konig, S. Hussain, B. Zupan, T. Curk, M. Frye, and J. Ule. 2012. 'Analysis of CLIP and iCLIP methods for nucleotide-resolution studies of protein-RNA interactions', *Genome Biol*, 13: R67.

Sugimoto, Y., A. Vigilante, E. Darbo, A. Zirra, C. Militti, A. D'Ambrogio, N. M. Luscombe, and J. Ule. 2015. 'hiCLIP reveals the in vivo atlas of mRNA secondary structures recognized by Staufen 1', *Nature*, 519: 491-4.

Sun, J. S., and J. L. Manley. 1995. 'A novel U2-U6 snRNA structure is necessary for mammalian mRNA splicing', *Genes Dev*, 9: 843-54.

Tagwerker, C., K. Flick, M. Cui, C. Guerrero, Y. Dou, B. Auer, P. Baldi, L. Huang, and P. Kaiser. 2006. 'A tandem affinity tag for two-step purification under fully denaturing conditions: application in ubiquitin profiling and protein complex identification combined with in vivocross-linking', *Mol Cell Proteomics*, 5: 737-48.

Tanaka, N., and B. Schwer. 2005. 'Characterization of the NTPase, RNA-binding, and RNA helicase activities of the DEAH-box splicing factor Prp22', *Biochemistry*, 44: 9795-803.

Tanner, N. K. 2003. 'The newly identified Q motif of DEAD box helicases is involved in adenine recognition', *Cell Cycle*, 2: 18-9.

Tardiff, D. F., and M. Rosbash. 2006. 'Arrested yeast splicing complexes indicate stepwise snRNP recruitment during in vivo spliceosome assembly', *RNA (New York, N.Y.)*, 12: 968-79.

Tarn, W. Y., K. R. Lee, and S. C. Cheng. 1993. 'The yeast PRP19 protein is not tightly associated with small nuclear RNAs, but appears to associate with the spliceosome after binding of U2 to the pre-mRNA and prior to formation of the functional spliceosome', *Mol Cell Biol*, 13: 1883-91.

Tauchert, M. J., J. B. Fourmann, R. Luhrmann, and R. Ficner. 2017. 'Structural insights into the mechanism of the DEAH-box RNA helicase Prp43', *Elife*, 6.

Theuser, M., C. Hobartner, M. C. Wahl, and K. F. Santos. 2016. 'Substrate-assisted mechanism of RNP disruption by the spliceosomal Brr2 RNA helicase', *Proc Natl Acad Sci U S A*, 113: 7798-803.

Tourigny, D. S., I. S. Fernandez, A. C. Kelley, and V. Ramakrishnan. 2013. 'Elongation factor G bound to the ribosome in an intermediate state of translocation', *Science*, 340: 1235490.

Tseng, C. K., and S. C. Cheng. 2008. 'Both catalytic steps of nuclear pre-mRNA splicing are reversible', *Science*, 320: 1782-4.

Tseng, C. K., H. L. Liu, and S. C. Cheng. 2011. 'DEAH-box ATPase Prp16 has dual roles in remodeling of the spliceosome in catalytic steps', *RNA (New York, N.Y.)*, 17: 145-54.

Valencia, P., A. P. Dias, and R. Reed. 2008. 'Splicing promotes rapid and efficient mRNA export in mammalian cells', *Proc Natl Acad Sci U S A*, 105: 3386-91.

van der Feltz, C., A. C. DeHaven, and A. A. Hoskins. 2018. 'Stress-induced Pseudouridylation Alters the Structural Equilibrium of Yeast U2 snRNA Stem II', *J Mol Biol*, 430: 524-36.

van der Feltz, C., and A. A. Hoskins. 2017. 'Methodologies for studying the spliceosome's RNA dynamics with single-molecule FRET', *Methods*, 125: 45-54.

Van Nostrand, Eric L., Gabriel A. Pratt, Alexander A. Shishkin, Chelsea Gelboin-Burkhart, Mark Y. Fang, Balaji Sundararaman, Steven M. Blue, Thai B. Nguyen, Christine Surka, Keri Elkins, Rebecca Stanton, Frank Rigo, Mitchell Guttman, and Gene W. Yeo. 2016. 'Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced CLIP (eCLIP)', *Nature Methods*, 13: 508.

van Nues, R. W., and J. D. Beggs. 2001. 'Functional contacts with a range of splicing proteins suggest a central role for Brr2p in the dynamic control of the order of events in spliceosomes of Saccharomyces cerevisiae', *Genetics*, 157: 1451-67.

Vetter, I. R., and A. Wittinghofer. 2001. 'The guanine nucleotide-binding switch in three dimensions', *Science*, 294: 1299-304.

Vijayraghavan, U., R. Parker, J. Tamm, Y. Iimura, J. Rossi, J. Abelson, and C. Guthrie. 1986. 'Mutations in conserved intron sequences affect multiple steps in the yeast splicing pathway, particularly assembly of the spliceosome', *Embo J*, 5: 1683-95.

Wagner, J. D., E. Jankowsky, M. Company, A. M. Pyle, and J. N. Abelson. 1998. 'The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes', *Embo J*, 17: 2926-37.

Wahl, M. C., C. L. Will, and R. Luhrmann. 2009. 'The spliceosome: design principles of a dynamic RNP machine', *Cell*, 136: 701-18.

Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay. 1982. 'Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold', *Embo J*, 1: 945-51.

Wallace, E. W. J., and J. D. Beggs. 2017. 'Extremely fast and incredibly close: cotranscriptional splicing in budding yeast', *RNA (New York, N.Y.)*, 23: 601-10.

Wan, R., C. Yan, R. Bai, G. Huang, and Y. Shi. 2016. 'Structure of a yeast catalytic step I spliceosome at 3.4 Å resolution', *Science*, 353: 895-904.

Wan, R., C. Yan, R. Bai, J. Lei, and Y. Shi. 2017. 'Structure of an Intron Lariat Spliceosome from Saccharomyces cerevisiae', *Cell*, 171: 120-32.e12.

Warkocki, Z., P. Odenwalder, J. Schmitzova, F. Platzmann, H. Stark, H. Urlaub, R. Ficner, P. Fabrizio, and R. Luhrmann. 2009. 'Reconstitution of both steps of Saccharomyces cerevisiae splicing with purified spliceosomal components', *Nat Struct Mol Biol*, 16: 1237-43.

Wilkinson, M. E., S. M. Fica, W. P. Galej, C. M. Norman, A. J. Newman, and K. Nagai. 2017. 'Postcatalytic spliceosome structure reveals mechanism of 3'-splice site selection', *Science*, 358: 1283-88.

Wilkinson, M. E., P. C. Lin, C. Plaschka, and K. Nagai. 2018. 'Cryo-EM Studies of Pre-mRNA Splicing: From Sample Preparation to Model Visualization', *Annu Rev Biophys*, 47: 175-99.

Will, C. L., and R. Luhrmann. 2011. 'Spliceosome structure and function', *Cold Spring Harb Perspect Biol*, 3.

Wittinghofer, A., and I. R. Vetter. 2011. 'Structure-function relationships of the G domain, a canonical switch motif', *Annu Rev Biochem*, 80: 943-71.

Wlodaver, A. M., and J. P. Staley. 2014. 'The DExD/H-box ATPase Prp2p destabilizes and proofreads the catalytic RNA core of the spliceosome', *RNA (New York, N.Y.)*, 20: 282-94.

Wu, C. G., and M. Spies. 2013. 'Overview: what are helicases?', *Adv Exp Med Biol*, 767: 1-16.

Xu, Y. Z., and C. C. Query. 2007. 'Competition between the ATPase Prp5 and branch region-U2 snRNA pairing modulates the fidelity of spliceosome assembly', *Mol Cell*, 28: 838-49.

Yan, C., J. Hang, R. Wan, M. Huang, C. C. Wong, and Y. Shi. 2015. 'Structure of a yeast spliceosome at 3.6-angstrom resolution', *Science*, 349: 1182-91.

Yan, Chuangye, Ruixue Wan, Rui Bai, Gaoxingyu Huang, and Yigong Shi. 2016a. 'Structure of a yeast activated spliceosome at 3.5 Å resolution', *Science*, 353: 904-11.

———. 2016b. 'Structure of a yeast step II catalytically activated spliceosome', *Science*.

Yang, F., X. Y. Wang, Z. M. Zhang, J. Pu, Y. J. Fan, J. Zhou, C. C. Query, and Y. Z. Xu. 2013. 'Splicing proofreading at 5' splice sites by ATPase Prp28p', *Nucleic Acids Res*, 41: 4660-70.

Yang, Q., and E. Jankowsky. 2005. 'ATP- and ADP-dependent modulation of RNA unwinding and strand annealing activities by the DEAD-box protein DED1', *Biochemistry*, 44: 13591-601.

Yeh, T. C., H. L. Liu, C. S. Chung, N. Y. Wu, Y. C. Liu, and S. C. Cheng. 2011. 'Splicing factor Cwc22 is required for the function of Prp2 and for the spliceosome to escape from a futile pathway', *Mol Cell Biol*, 31: 43-53.

Zarnegar, Brian J., Ryan A. Flynn, Ying Shen, Brian T. Do, Howard Y. Chang, and Paul A. Khavari. 2016. 'irCLIP platform for efficient characterization of protein–RNA interactions', *Nature Methods*, 13: 489.

Zhang, M., and M. R. Green. 2001. 'Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56', *Genes Dev*, 15: 30-5.

Zhou, Z., L. J. Licklider, S. P. Gygi, and R. Reed. 2002. 'Comprehensive proteomic analysis of the human spliceosome', *Nature*, 419: 182-5.

Zhou, Z., J. Sim, J. Griffith, and R. Reed. 2002. 'Purification and electron microscopic visualization of functional human spliceosomes', *Proc Natl Acad Sci U S A*, 99: 12203-7.

Zubradt, M., P. Gupta, S. Persad, A. M. Lambowitz, J. S. Weissman, and S. Rouskin. 2017. 'DMS-MaPseq for genome-wide or targeted RNA structure probing in vivo', *Nat Methods*, 14: 75-82.

8 Appendix

The NGS data which is discussed in this thesis can be found under the link listed below. Bedgraph-files (all discussed experiments) and bam-files (session_12 and 13) visualise the crosslink sites and the sequencing coverage, respectively, when opened in the IGV viewer. Therefore, the respective substrate-ome or the yeast genome (sacCer3) has to be loaded. Files named 'IGV_session_*xxx*' in the respective folder list the respective genome; substrate-omes are kept in the 'substrate-omes'-folder. The pipeline written in Snakemake, which contains the processing code can equally be found in the linked folder.

Linked folder: https://cloud.mrc-lmb.cam.ac.uk/index.php/s/4xNtylyhtjeuJAE
Password: isCL2018

The appendix contains additional statistical data on *in vitro* spliceosome iCLIP data, such as the mapping distribution between unmapped reads, genome and substrate-ome mapped reads (**Figure 8-1** and **Figure 8-2**) and the length distribution of cDNA reads (**Figure 8-3**). Additionally, the appendix contains two tables, which lists constructs for the protein expression described in chapter 2 (**Table 8-1** and **Table 8-2**).

Figure 8-1: Read distribution for control experiments, Prp22 and SmB *in vitro* spliceosome iCLIP data. The proportion and number of reads, which remained unmapped (green), or map to the genome (blue) and substrate-ome (navy) are compared before PCR duplicates were removed. Grey bars show the proportion of PCR duplicates calculated based on reads which map the substrate-ome. Data indicated with a * derived from short-iCLIP experiments. The data was analysed and illustrated by Charlotte Capitanchik.

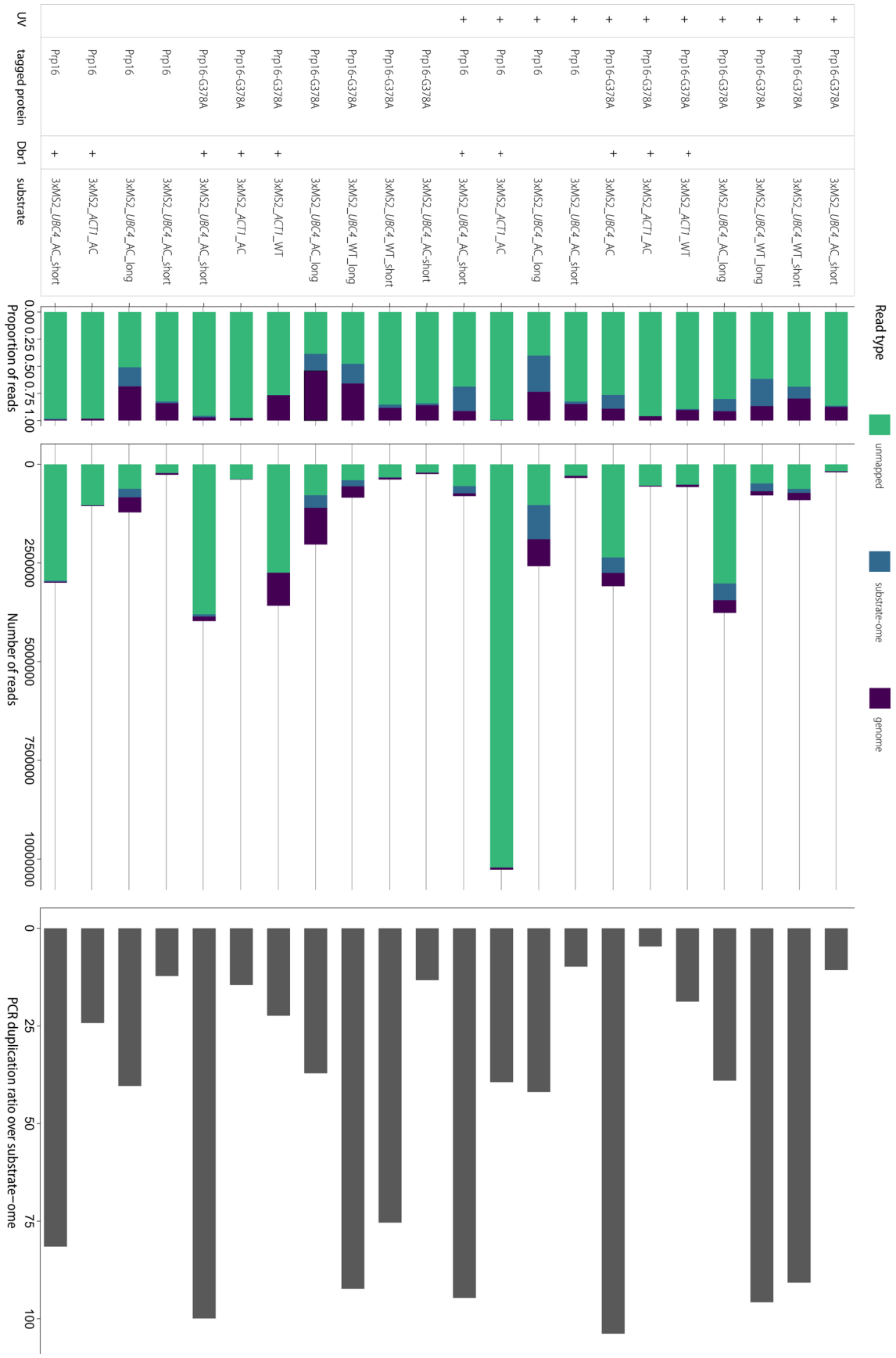


Figure 8-2: Read distribution for Prp16 *in vitro* spliceosome iCLIP data. The proportion and number of reads, which remained unmapped (green), or map to the genome (blue) and substrate-ome (navy) are compared before PCR duplicates were removed. Grey bars show the proportion of PCR duplicates calculated based on reads which map the substrate-ome. The data was analysed and illustrated by Charlotte Capitanchik.

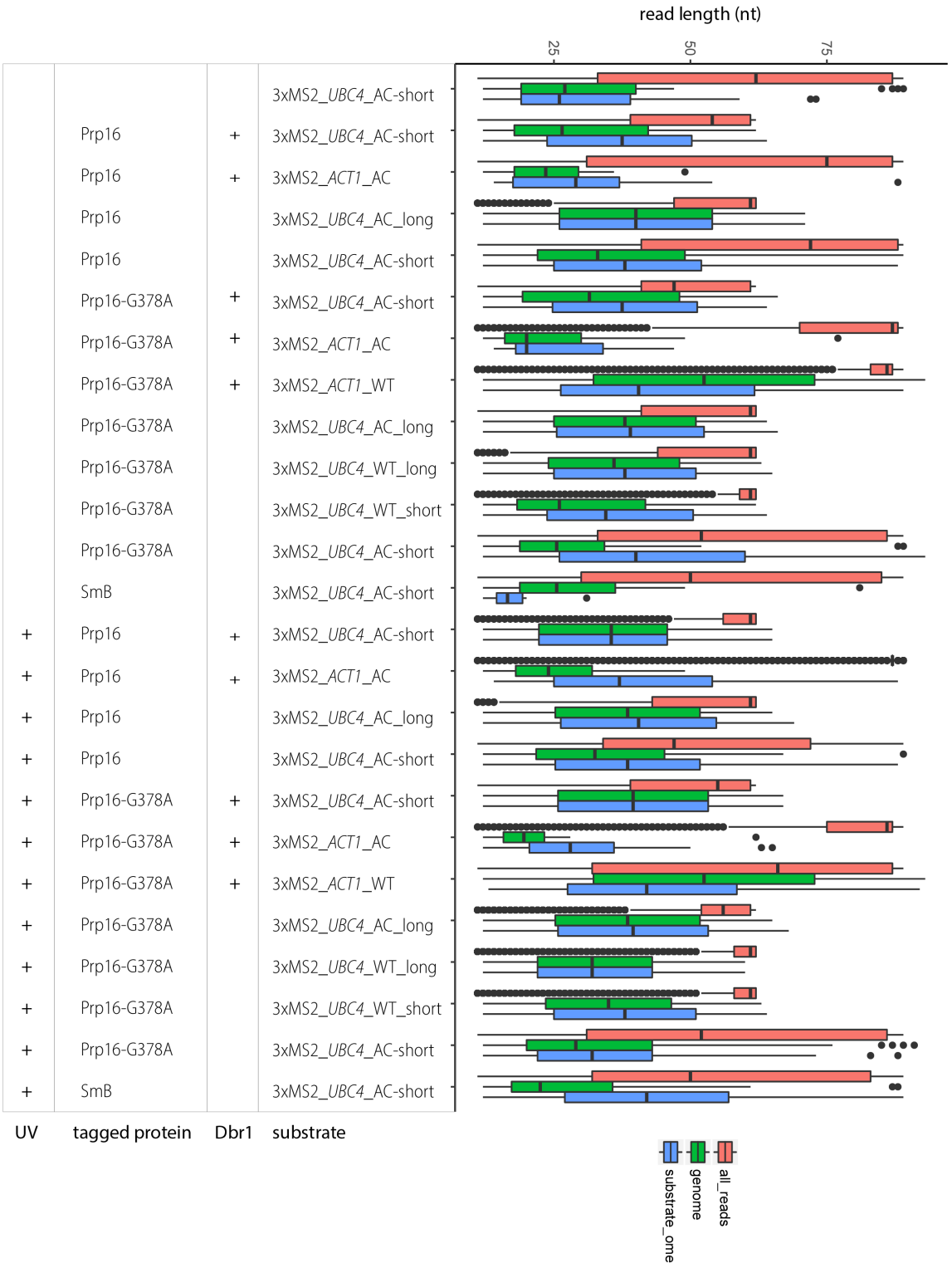


Figure 8-3: cDNA length distribution of *in vitro* spliceosome iCLIP data. The read distribution of all reads (red) was determined by the length of all reads before mapping but removing PCR duplicates. For the distribution of reads mapping to the yeast genome (green) and substrate-ome (blue), read length were taken from the mapped bed-files and PCR duplicates removed. Reads spanning splice sites were excluded for genome-mappers. Charlotte Capitanchik wrote an R script to produce the resented box plot. The median is indicated with a black bar, outliers represented by black dots.

Table 8-1: Overview of expression tests for Prp8-N. Constructs were either expressed in bacteria or yeast (column 1). The variation in tag/protease site and the amino acid fragment of the construct are given in column 2 and 3, respectively. The last column indicates under which conditions the expression was successful. Constructs indicated with a * are shown in purification in Figure 2-2.

expression system (vector)	tag / protease site	amino acid fragment (add. TEV cleavage site)	success
Bacterial (pGEX-6P-I)	N-GST-Pre-TEV	22-411, 69-411, 22-543, 69-543, 412-770, 540-770, 22-770, 69-770	Expression low
Bacterial (pET28a)	N-His6-TEV	22-411, 69-411, 22-543, 69-543, 412-770, 540-770, 22-770, 69-770	Expression low
Bacterial (pETM-11)	N-His6-TEV	22-411, 69-411, 22-543, 69-543, 412-770, 540-770, 22-770, 69-770	Expression low
Bacterial (pETM-41)	N-MBP-His6-TEV	22-411, 69-411, 22-543, 69-543, 412-770, 540-770, 22-770, 69-770	Expression low
Bacterial (pETM-40)	N-MBP-TEV	22-411, 69-411, 22-543, 69-543, 412-770, 540-770, 22-770, 69-770	Expression low
Bacterial codon-optimised (pETM-41)	N-MBP-His6-TEV	1-770	Low protein yield
Bacterial codon-optimised (pETM-40)	N-MBP-His6	1-770	Low protein yield
Yeast (pRS426)	N-CBP-His8-TEV	1-770	Good co-expression with Snu114*
Yeast (pRS426)	N-CBP-His8	1-411 TEV(122)	No expression
Yeast (pRS426)	N-CBP-His8	1-543 TEV(122)	No expression
Yeast (pRS426)	N-CBP-His8	1-735 TEV(122)	Bad expression
Yeast (pRS426)	N-CBP-His8	1-770 TEV(122)	Good co-expression with Snu114, TEV accessible
Yeast (pRS426)	N-CBP-His8-TEV	1-770 TEV(328)	Good co-expression with Snu114, TEV site inaccessible
Yeast (pRS426)	N-CBP-His8	1-770 TEV(328)	Good co-expression with Snu114, TEV site inaccessible
Yeast (pRS426)	N-CBP-His8-TEV	1-770 TEV(380)	Good co-expression with Snu114, TEV site inaccessible
Yeast (pRS426)	N-CBP-His8-TEV	1-770 TEV(318)	Good co-expression with Snu114, cleaved fragment not released
Yeast (pRS426)	N-CBP-His8-TEV	1-770 spacer_TEV(328)	Good co-expression with Snu114, cleaved fragment not released
Yeast (pRS426)	N-CBP-His8-TEV	1-770 spacer_TEV(380)	Good co-expression with Snu114, cleaved fragment not released

Table 8-2: Overview of expression tests for Snu114. Constructs were either expressed in bacteria or yeast (column 1). The variation in tag/protease site and the amino acid fragment of the construct are given in column 2 and 3, respectively. The last column indicates under which conditions the expression was successful. Constructs indicated with a * are shown in purification in Figure 2-2.

expression system (vector)	**tag / protease site**	**amino acid fragment (add. TEV site)**	**success**
Yeast (pRS424)	TEV-His8-C	Snu114 (full length)	Incomplete cleavage
Yeast (pRS424)	-	Snu114 (full length)	Good expression*
Yeast (pRS424)	-	Snu114 (full length) TEV(70)	Good expression